



3. Document Type Definition

Branislava Šandrih

branislava.sandrih@fil.bg.ac.rs

NAPOMENA: Sadržaj ove prezentacije preuzet je od prof. Cvetane Krstev sa
<http://poincare.matf.bg.ac.rs/~cvetana/kurs-xml/>

DTD

- Document Type Definition
 - Definicija tipa dokumenta
- XML dokumenti mogu da budu veoma fleksibilni
- Sve aplikacije ne mogu da podrže takvu fleksibilnost, odnosno, ona je za njih nepoželjna
 - Na primer, u XHTML-u element `<li>` može da bude dete samo od `<ol>` i `<ul>`, dok njegovo pojavljivanje direktno u sadržaju drugih elemenata, recimo `<p>` ili `<blockquote>` zbunjuje pregledač
- Uloga DTD-a je da **precizno, koristeći se formalnom sintaksom**, opiše koji se elementi mogu pojaviti u dokumentu, šta je mogući sadržaj svih dozvoljenih elemenata, šta su njihovi atributi i njihove dozvoljene vrednosti

Validni XML dokumenti

- XML dokument koji koristi DTD je **validan** ako je zapisan u skladu sa DTD-em koji dokument koristi
- Validnost proverava XML parser
- Ove greške se u nekim slučajevima ne moraju smatrati fatalnim, ali to zavisi od aplikacije. Greška u validnost može biti fatalna za slogove kojima se puni baza podataka (na primer, ako nedostaje obavezno polje), dok pregledač ovakvu grešku može da zaobiđe
 - drugim rečima, **XML dokument mora da bude dobro formiran ali ne mora da bude validan**
- DTD još uvek ostavlja dosta fleksibilnosti autorima. Na primer, on ne govori:
 - šta je koreni element dokumenta
 - koliko se mnogo primeraka nekog elementa može pojaviti u dokumentu
 - kako izgledaju karakterski podaci unutar nekog elementa
 - šta je značenje pojedinačnih elemenata (na primer, da li element sadrži ime osobe ili datum)

DTD primer

```
<!-- DTD za element *osoba* -->  
<!ELEMENT osoba (ime,profesija*)>  
<!ELEMENT ime (lično_ime, prezime)>  
<!ELEMENT lično_ime (#PCDATA)>  
<!ELEMENT prezime (#PCDATA)>  
<!ELEMENT profesija (#PCDATA)>
```

```
<!-- Primer validnog dokumenta -->  
<osoba>  
  <ime>  
    <lično_ime>Alan</lično_ime>  
    <prezime>Tjuring</prezime>  
  </ime>  
</osoba>
```

```
<!-- Primer nevalidnog dokumenta -->  
<osoba>  
  <ime>  
    <prezime>Tjuring</prezime>  
    <lično_ime>Alan</lično_ime>  
  </ime>  
</osoba>
```

Primeri iz HTML DTD-a

<!-- deklaracija uređene liste, element *UL* -->

<!ELEMENT UL (LI)+>

<!-- deklaracija liste definicija, element *DL* -->

<!ELEMENT DL (DT|DD)+>

<!-- deklaracija horizontalne linije, element *HR* -->

<!ELEMENT HR EMPTY>

<!-- deklaracija tabele, element *TABLE* -->

<!ELEMENT TABLE (CAPTION?, TR+)>

<!ELEMENT TR (TH|TD)*>

<!-- deklaracija elementa za selekciju više mogućnosti, element *SELECT* -->

<!ELEMENT SELECT (OPTION+)>

<!ELEMENT OPTION (#PCDATA)>

Referisanje na spoljašnji DTD

- Validni dokument referiše na DTD u odnosu na koji se njegova validnost proverava
- Ova referenca se zadaje u jedinstvenoj **deklaraciji tipa dokumenta** koja govori šta je koreni element dokumenta i na kom URL-u se DTD može naći
- Ova deklaracija se smešta u **prolog** dokumenta, a to je deo XML dokumenta između XML deklaracije i početne etikete korenog elementa

Primer referisanja na spoljni DTD

```
<?xml version="1.0" encoding="ISO-8859-2" standalone="no"?>
<!DOCTYPE osoba SYSTEM "http://nlp.matf.bg.ac.rs/xml/dtds/osoba.dtd">
<osoba>
  <ime>
    <lično_ime>Alan</lično_ime>
    <prezime>Tjuring</prezime>
  </ime>
  <profesija>informatičar</profesija>
  <profesija>matematičar</profesija>
  <profesija>kriptograf</profesija>
</osoba>
```

Referisanje na unutrašnji DTD

- Umesto da se referiše preko URL-a, DTD može da bude sadržan u deklaraciji tipa dokumenta unutar uglastih zagrada

```
<?xml version="1.0" encoding="ISO-8859-2" standalone="yes"?>
<!DOCTYPE osoba [
  <!-- DTD za element *osoba* -->
  <!ELEMENT osoba (ime,profesija*)>
  <!ELEMENT ime (lično_ime, prezime)>
  <!ELEMENT lično_ime (#PCDATA)>
  <!ELEMENT prezime (#PCDATA)>
  <!ELEMENT profesija (#PCDATA)>
]>
<osoba>
  <ime>
    <lično_ime>Alan</lično_ime>
    <prezime>Tjuring</prezime>
  </ime>
  <profesija>informatičar</profesija>
  <profesija>matematičar</profesija>
  <profesija>kriptograf</profesija>
</osoba>
```


Kombinovanje unutrašnjih i spoljašnjih DTD

- Ponekad deklaracija tipa dokumenta sadrži neke deklaracije direktno a ostale referiše preko URL-a
- Ta dva podskupa deklaracija moraju da budu **kompatibilna**, što znači da ni jedan od njih ne može da poništi deklaraciju elementa ili atributa koju je uveo drugi podskup

```
<?xml version="1.0" encoding="ISO-8859-2" standalone="no"?>  
<!DOCTYPE osoba SYSTEM "ime.dtd" [  
  <!-- DTD za elemente *osoba* i *profesija*-->  
  <!ELEMENT profesija (#PCDATA)>  
  <!ELEMENT osoba (ime,profesija*)>  
>
```

Deklaracije elemenata

- Svaki element koji se koristi u validnom XML dokumentu mora da bude deklarisan u DTD-u. Deklaracija elementa ima oblik:

```
<!ELEMENT ime_elementa (model_sadržaja)>
```

- **Model sadržaja** specifikuje šta sve mogu, odnosno moraju, da budu deca nekog elementa i u kom redosledu se deca pojavljuju

Model sadržaja (1)

- **Parsirani karakterski podaci**

- sadržaj elementa su karakterski podaci, a ne drugi elementi. Na primer,

<!ELEMENT TEXTAREA (#PCDATA)>

- **Elementi deca**

- sadržaj elementa je jedan element dete određenog tipa. Na primer,

<!ELEMENT Fax (Tel_Broj)>

- **Nizovi elemenata deca**

- sadržaj elementa su elementi deca koji se moraju u primerku validnog dokumenta pojavljivati u **naznačenom redosledu**
 - elementi deca se u deklaraciji modela sadržaja razdvajaju zapetom. Na primer,

<!ELEMENT ime (lično_ime,prezime)>

Your α

Model sadržaja (2)

- **Izbor između više elemenata dece**

- ponekad element u jednom primerku dokumenta može da sadrži jedan element dete, a u drugom primerku dokumenta drugi element dete. Izbor između dva elementa se u modelu sadržaja naznačava vertikalnom crtom. Na primer,

<!ELEMENT naslov (h1 | h2 | h3)>

- **Broj dece**

- neki elementi deca su u sadržaju elementa opcioni, a neki se mogu pojaviti više puta
- imenu elementa deteta se u modelu sadržaja može dodati jedan od tri sufiksa:

- **?** Element je opcioni (nula ili jedno pojavljivanje)
- ***** Element je opcioni i ponovljiv (nula ili više pojavljivanja)
- **+** Element je obavezan i ponovljiv (jedno ili više pojavljivanja)

<!ELEMENT TABLE (CAPTION?,TR+)>

Model sadržaja (3)

- **Upotreba zagrada**

- nizovi, izbori i indikatori pojavljivanja mogu se kombinovati u jednom modelu sadržaja. Treba samo voditi računa da se unutar jednog para zagrada mogu koristiti samo nizovi ili samo izbori. Na primer,

<!-- Tačka u ravni zadaje se ili Dekartovima ili polarnim koordinatama -->
<!ELEMENT tačka ((x,y) | (r,u))>

- **Prazan sadržaj**

- neki elementi nemaju sadržaj. To su **prazni elementi**. Oni ne moraju imati završnu etiketu ako se početna etiketa završava sa />

- deklaracija praznog elementa je,

<!ELEMENT IMG EMPTY>

Model sadržaja (4)

- **Mešoviti sadržaj**

- sadržaj elementa <biografija> koji je korišćen za obeležavanje biografije Alana Tjuringa bio bi deklarisan na sledeći način:

```
<!ELEMENT biografija  
    (#PCDATA | ime | datum | istaknuto | profesija)*  
>
```

- Mešoviti sadržaj se uvek zadaje ovakvom vrstom deklaracije. Treba zapamtiti da:
 - broj elemenata dece u mešovitom sadržaju nekog elementa nije ograničen
 - parsirani karakterski podaci (#PCDATA) moraju uvek da budu **prvi** u listi koja deklariše mešoviti sadržaj
 - mešoviti sadržaj se nikada ne može deklarirati tako da sadrži tačno jedan element dete
 - mešoviti sadržaj se nikada ne može deklarirati tako da parsirani karakterski podaci dođu pre, ili posle, nekog elementa deteta
 - mešoviti sadržaj ne može da bude deo nekog složenijeg modela sadržaja

Model sadržaja (5)

- **Proizvoljan sadržaj**
 - U komotnom DTD-u, neki elementi mogu imati proizvoljan sadržaj, što znači da mogu imati mešoviti sadržaj u kome se kao deca pojavljuju ma koji elementi deklarirani u tom DTD-u, uključujući i sam element koji se deklarira
- Deklaracija jednog ovakvog elementa je
<!ELEMENT strana ANY>

Atributi (1)

- U validnom XML dokumentu moraju da budu deklarirani, pored svih elemenata koji se koriste, i svi atributi tih elemenata
- Atributi se ne deklariraju nezavisno od elemenata
- Ako je neki atribut svojstvo više elemenata, on se mora deklarirati za svaki od tih elemenata
- Deklaracija atributa ima oblik:
`<!ATTLIST ime_elementa ime_atributa deklarirana_vred podrazumevana_vred>`

Atributi (2)

- Primer jedne deklaracije atributa u HTML DTD je:

```
<!ELEMENT H1 (%text;)*>
```

```
<!ATTLIST H1 align (left|center|right) #IMPLIED>
```

- U okviru jedne deklaracije atributa može se deklarirati više atributa jednog elementa. Na primer,

```
<!ELEMENT HR EMPTY>
```

```
<!ATTLIST HR
```

```
align (left|center|right) #IMPLIED
```

```
noshade (noshade) #IMPLIED
```

```
size %Pixels; #IMPLIED
```

```
width %Length; #IMPLIED>
```

Tipovi atributa (1)

- U dobro formiranom XML dokumentu vrednost atributa može da bude proizvoljna niska karaktera
- DTD dozvoljava da se postave dodatna ograničenja na vrednosti atributa
- Šta više, ta ograničenja su **jača** od onih koja se mogu postaviti na sadržaje elemenata
- U XML-u postoji deset tipova atributa. To su:
 - **CDATA** (character data)
 - **<!ELEMENT IMG EMPTY>**
 - **<!ATTLIST IMG ALT CDATA #IMPLIED>**
 - **NMTOKEN** (name token)
 - vrednost atributa je legalno XML ime, s tim da se prvi karakter imena ne ograničava na slovo i podvlaku
 - **<!ELEMENT META EMPTY>**
 - **<!ATTLIST META NAME NMTOKEN #IMPLIED>**
 - **NMTOKENS** (name tokens)
 - vrednost atributa je lista XML imena razdvojenih belinama

Tipovi atributa (2)

- Nabrajanje

- ovo je jedini XML atribut koji nije XML ključna reč
- navode se sve dozvoljene vrednosti atributa razdvojene vertikalnom crtom
- svaka dozvoljena vrednost mora da bude XML ime

<!ELEMENT UL (LI)+>

<!ATTLIST UL TYPE (disc|square|circle) #IMPLIED>

- ID (identifier)

- vrednost atributa ovog tipa mora da bude XML ime koje je jedinstveno u okviru dokumenta
- element može da ima samo jedan atribut ovog tipa

<!ATTLIST student broj_indeksa ID #REQUIRED>

<!ATTLIST predmet šifra ID #REQUIRED>

- IDREF (identifier reference)

- na vrednost ovog atributa referiše atribut tipa ID nekog elementa
- prema tome i on mora da bude XML ime
- korišćenje ovog atributa omogućava uspostavljanje mnogo-mnogo veza među elementima

<!ATTLIST sluša_predmet predmet IDREF #REQUIRED>

<!ATTLIST slušalac_predmeta student IDREF #REQUIRED>

Tipovi atributa (3)

- IDREFS (identifier references)

- vrednost ovog atributa je lista XML imena razdvojenih belinama, od kojih je svako ime ID nekog elementa iz istog dokumenta
- u okviru jedne deklaracije atributa može se deklarirati više atributa jednog elementa

```
<!ATTLIST student broj_indeksa ID #REQUIRED  
    sluša_predmete IDREFS #REQUIRED>  
<!ATTLIST predmet šifra ID #REQUIRED  
    slušaoci_predmeta IDREFS #REQUIRED>
```

- ENTITY

- vrednost atributa ovog tipa je ime nekog neparsiranog entiteta koji deklarisan u istom DTD-u

- ENTITIES

- vrednost atributa ovog tipa je lista imena neparsiranih entiteta koji su deklarirani u istom DTD-u
- imena se u listi razdvajaju belinama

- NOTATION

- vrednost atributa ovog tipa je ime notacije koja je deklarirana u istom DTD-u

Podrazumevane vrednosti atributa

- U ATTLIST deklaraciji atributa se osim njegovog tipa navodi i njegova podrazumevana vrednost.
- Postoje četiri mogućnosti:
 - **#IMPLIED** Atribut nije obavezan (opcionalan), neki primerci elemenata ga mogu imati ali ne moraju svi. Podrazumevana vrednost ne postoji
 - **#REQUIRED** Atribut je obavezan i svaki primerak elementa ga mora imati. Podrazumevana vrednost ne postoji
 - **#FIXED** Vrednost atributa je konstantna i nepromenljiva, bez obzira da li je atribut naveden u primerku elementa ili ne
 - **Navedena vrednost** Podrazumevana vrednost je ona koja je navedena pod navodnicima

Your company name

Karakterski entiteti (1)

- Karakterski entiteti omogućavaju korišćenje karaktera kojih nema u karakterskom skupu koji se koristi, ili kojih nema na tastaturi
- Karakterski entiteti se ne deklarišu, oni se jednostavno koriste
 - iza znaka **&#** navodi se pozicija karaktera u karakterskom skupu koji se koristi.
 - na primer, ako se koristi karakterski skup **ISO 8859-1 Latin 1**, onda je **¡** okrenuti znak uzvika (¡), jer je 161 pozicija tog karaktera u tom karakterskom skupu
 - slično je **£** znak za funtu sterlinga (£) jer je 163 pozicija znaka za funtu u istom karakterskom skupu.

intermezzo: ASCII

ASCII TABLE

Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char
0	0	[NULL]	32	20	[SPACE]	64	40	@	96	60	`
1	1	[START OF HEADING]	33	21	!	65	41	A	97	61	a
2	2	[START OF TEXT]	34	22	"	66	42	B	98	62	b
3	3	[END OF TEXT]	35	23	#	67	43	C	99	63	c
4	4	[END OF TRANSMISSION]	36	24	\$	68	44	D	100	64	d
5	5	[ENQUIRY]	37	25	%	69	45	E	101	65	e
6	6	[ACKNOWLEDGE]	38	26	&	70	46	F	102	66	f
7	7	[BELL]	39	27	'	71	47	G	103	67	g
8	8	[BACKSPACE]	40	28	(	72	48	H	104	68	h
9	9	[HORIZONTAL TAB]	41	29	)	73	49	I	105	69	i
10	A	[LINE FEED]	42	2A	*	74	4A	J	106	6A	j
11	B	[VERTICAL TAB]	43	2B	+	75	4B	K	107	6B	k
12	C	[FORM FEED]	44	2C	,	76	4C	L	108	6C	l
13	D	[CARRIAGE RETURN]	45	2D	-	77	4D	M	109	6D	m
14	E	[SHIFT OUT]	46	2E	.	78	4E	N	110	6E	n
15	F	[SHIFT IN]	47	2F	/	79	4F	O	111	6F	o
16	10	[DATA LINK ESCAPE]	48	30	0	80	50	P	112	70	p
17	11	[DEVICE CONTROL 1]	49	31	1	81	51	Q	113	71	q
18	12	[DEVICE CONTROL 2]	50	32	2	82	52	R	114	72	r
19	13	[DEVICE CONTROL 3]	51	33	3	83	53	S	115	73	s
20	14	[DEVICE CONTROL 4]	52	34	4	84	54	T	116	74	t
21	15	[NEGATIVE ACKNOWLEDGE]	53	35	5	85	55	U	117	75	u
22	16	[SYNCHRONOUS IDLE]	54	36	6	86	56	V	118	76	v
23	17	[ENG OF TRANS. BLOCK]	55	37	7	87	57	W	119	77	w
24	18	[CANCEL]	56	38	8	88	58	X	120	78	x
25	19	[END OF MEDIUM]	57	39	9	89	59	Y	121	79	y
26	1A	[SUBSTITUTE]	58	3A	:	90	5A	Z	122	7A	z
27	1B	[ESCAPE]	59	3B	;	91	5B	[	123	7B	{
28	1C	[FILE SEPARATOR]	60	3C	<	92	5C	\	124	7C	
29	1D	[GROUP SEPARATOR]	61	3D	=	93	5D	]	125	7D	}
30	1E	[RECORD SEPARATOR]	62	3E	>	94	5E	^	126	7E	~
31	1F	[UNIT SEPARATOR]	63	3F	?	95	5F	_	127	7F	[DEL]

intermezzo: ISO-8859-1

	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F
00	<u>NUL</u> 0000	<u>STX</u> 0001	<u>SOT</u> 0002	<u>ETX</u> 0003	<u>EOT</u> 0004	<u>ENQ</u> 0005	<u>ACK</u> 0006	<u>BEL</u> 0007	<u>BS</u> 0008	<u>HT</u> 0009	<u>LF</u> 000A	<u>VT</u> 000B	<u>FF</u> 000C	<u>CR</u> 000D	<u>SO</u> 000E	<u>SI</u> 000F
10	<u>DLE</u> 0010	<u>DC1</u> 0011	<u>DC2</u> 0012	<u>DC3</u> 0013	<u>DC4</u> 0014	<u>NAK</u> 0015	<u>SYN</u> 0016	<u>ETB</u> 0017	<u>CAN</u> 0018	<u>EM</u> 0019	<u>SUB</u> 001A	<u>ESC</u> 001B	<u>FS</u> 001C	<u>GS</u> 001D	<u>RS</u> 001E	<u>US</u> 001F
20	<u>SP</u> 0020	!	"	#	\$	%	&	'	(	)	*	+	,	-	.	/
30	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
40	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
50	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_
60	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
70	p	q	r	s	t	u	v	w	x	y	z	{		}	~	<u>DEL</u> 007F
80																
90																
A0	<u>NSBP</u> 00A0	¡	¢	£	¤	¥	¦	§	¨	©	ª	«	¬	­	®	¯
B0	°	±	²	³	´	µ	¶	·	¸	¹	º	»	¼	½	¾	¿
C0	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï
D0	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ	ß
E0	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï
F0	ð	ñ	ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ý	þ	ÿ

Your company

intermezzo: ISO-8859-2

	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F
00	<u>NUL</u> 0000	<u>STX</u> 0001	<u>SOT</u> 0002	<u>ETX</u> 0003	<u>EOT</u> 0004	<u>ENQ</u> 0005	<u>ACK</u> 0006	<u>BEL</u> 0007	<u>BS</u> 0008	<u>HT</u> 0009	<u>LF</u> 000A	<u>VT</u> 000B	<u>FF</u> 000C	<u>CR</u> 000D	<u>SO</u> 000E	<u>SI</u> 000F
10	<u>DLE</u> 0010	<u>DC1</u> 0011	<u>DC2</u> 0012	<u>DC3</u> 0013	<u>DC4</u> 0014	<u>NAK</u> 0015	<u>SYN</u> 0016	<u>ETB</u> 0017	<u>CAN</u> 0018	<u>EM</u> 0019	<u>SUB</u> 001A	<u>ESC</u> 001B	<u>FS</u> 001C	<u>GS</u> 001D	<u>RS</u> 001E	<u>US</u> 001F
20	<u>SP</u> 0020	!	"	#	\$	%	&	'	(	)	*	+	,	-	.	/
30	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
40	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
50	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_
60	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
70	p	q	r	s	t	u	v	w	x	y	z	{		}	~	<u>DEL</u> 007F
80																
90																
A0	<u>NSBP</u> 00A0	Ā 0104	ā 02D8	Ł 0141	* 00A4	Ĺ 013D	Š 015A	Š 00A7	ˆ 00A8	š 0160	š 015E	ť 0164	ž 0179	- 00AD	ž 017D	ž 017B
B0	° 00B0	ą 0105	ć 02DB	ł 0142	ł 00B4	ł 013E	ś 015B	ˆ 02C7	ˆ 00B8	š 0161	š 015F	ť 0165	ž 017A	ˆ 02DD	ž 017E	ž 017C
C0	Ř 0154	Á 00C1	Ā 00C2	Ā 0102	Ā 00C4	Ĺ 0139	Ć 0106	Ç 00C7	Č 010C	É 00C9	Ě 0118	Ě 00CB	Ě 011A	Í 00CD	Ī 00CE	Ď 010E
D0	Đ 0110	Ń 0143	Ń 0147	Ó 00D3	Ō 00D4	Ō 0150	Ö 00D6	× 00D7	Ř 0158	Ů 016E	Ú 00DA	Ú 0170	Ú 00DC	Ý 00DD	Ť 0162	ß 00DF
E0	í 0155	á 00E1	ā 00E2	ă 0103	ă 00E4	Ĺ 013A	Ć 0107	Ç 00E7	Č 010D	é 00E9	ę 0119	ě 00EB	ě 011B	í 00ED	î 00EE	ď 010F
F0	đ 0111	ń 0144	ň 0148	ó 00F3	ô 00F4	ő 0151	ö 00F6	÷ 00F7	ř 0159	ů 016F	ú 00FA	ú 0171	ù 00FC	ý 00FD	ț 0163	· 02D9

Your company

intermezzo: ISO-8859-5

	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F
00	<u>NUL</u> 0000	<u>STX</u> 0001	<u>SOT</u> 0002	<u>ETX</u> 0003	<u>EOT</u> 0004	<u>ENQ</u> 0005	<u>ACK</u> 0006	<u>BEL</u> 0007	<u>BS</u> 0008	<u>HT</u> 0009	<u>LF</u> 000A	<u>VT</u> 000B	<u>FF</u> 000C	<u>CR</u> 000D	<u>SO</u> 000E	<u>SI</u> 000F
10	<u>DLE</u> 0010	<u>DC1</u> 0011	<u>DC2</u> 0012	<u>DC3</u> 0013	<u>DC4</u> 0014	<u>NAK</u> 0015	<u>SYN</u> 0016	<u>ETB</u> 0017	<u>CAN</u> 0018	<u>EM</u> 0019	<u>SUB</u> 001A	<u>ESC</u> 001B	<u>FS</u> 001C	<u>GS</u> 001D	<u>RS</u> 001E	<u>US</u> 001F
20	<u>SP</u> 0020	!	"	#	\$	%	&	'	(	)	*	+	,	-	.	/
30	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
40	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
50	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_
60	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
70	p	q	r	s	t	u	v	w	x	y	z	{		}	~	<u>DEL</u> 007F
80																
90																
A0	<u>NSBP</u> 00A0	Ё	Ђ	Ѓ	Є	Ѕ	І	Ї	Ј	Љ	Њ	Ћ	Ќ	–	Ў	Џ
B0	А	В	В	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О	П
C0	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Ь	Э	Ю	Я
D0	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п
E0	р	с	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь	э	ю	я
F0	№	ё	ђ	ѓ	є	ѕ	і	ї	ј	љ	њ	ћ	ќ	ѕ	ў	џ
	2116	0451	0452	0453	0454	0455	0456	0457	0458	0459	045A	045B	045C	00A7	045E	045F

Your company

Karakterski entiteti (2)

- Ukoliko već nisu, karakterski entiteti se mogu imenovati (radi lakše upotrebe):

<!ENTITY iexcl "¡">

<!ENTITY pound "£">

Videti:

<http://www.oasis-open.org/docbook/specs/wd-docbook-xmlocharent-0.3.html#entities>

- Karakterski entiteti definisani u jednom DTD-u, mogu se koristiti u XML dokumentima koji koriste taj DTD pomoću reference entiteta

Your code: **&lmeEntiteta;**

(na primer, **¡** ili **£**)



Parametarski entiteti (1)

- Njihova osnovna namena je da **olakšaju i sistematizuju deklaracije modela sadržaja elemenata i njihovih lista atributa**
- Deklaracija parametarskog entiteta je oblika:
<!ENTITY % ime_entiteta tekst_entiteta>
- Tekst entiteta je tekst koji posle parsiranja, odnosno interpretacije, zamenjuje referencu entiteta
- Parametarski entitet se može referisati isključivo **u okviru DTD-a** na sledeći način:
%ime_entiteta;
- Parametarski entitet može da bude i **spoljašnji**, što je posebno značajno za kompleksne DTD-e, koji se mogu razbiti u module, a korisnik preko referenci parametarskih entiteta učitava samo one koji su mu u određenom trenutku potrebni
- Deklaracija spoljašnjeg parametarskog entiteta je:
<!ENTITY % ime_entiteta SYSTEM URI_teksta_entiteta>

Parametarski entiteti (2)

- Parametarski entiteti se mogu redefinisati

- Primer:

```
<!-- Prva deklaracija parametarskog entiteta x.bibl-->
```

```
<!ENTITY % x.bibl "" >
```

```
<!ENTITY % m.bibl "%x.bibl; bibl | bibl.full | bibl.struc">
```

```
<!-- Druga deklaracija (redefinicija) parametarskog entiteta x.bibl-->
```

```
<!ENTITY % x.bibl " moja.bibl | " >
```

- Posle učitavanja obe deklaracije (npr, u različitim spoljašnjim DTD-evima), tekst zamene entiteta m.bibl biće:

```
" moja.bibl | bibl | bibl.full | bibl.struct "
```

Primer upotrebe parametarskih entiteta u HTML DTD (1)

- Deklaracija modela sadržaja

```
<!ENTITY % font
  "TT | I | B | U | STRIKE | BIG | SMALL | SUB | SUP">
<!ENTITY % phrase
  "EM | STRONG | DFN | CODE | SAMP | KBD | VAR | CITE">
<!ENTITY % special
  "A | IMG | APPLET | FONT | BASEFONT | BR | SCRIPT">
<!ENTITY % form
  "INPUT | SELECT | TEXTAREA">
<!ENTITY % text
  "#PCDATA | %font; | %phrase; | %special; | %form;">
```

```
<!ELEMENT TD (%text;)*>
```



```
<!ELEMENT TD (#PCDATA | TT | I | B | U | STRIKE | BIG | SMALL | SUB | SUP | EM | STRONG | DFN | CODE | SAMP |
KBD | VAR | CITE | A | IMG | APPLET | FONT | BASEFONT | BR | SCRIPT | INPUT | SELECT | TEXTAREA)*>
```

Primer upotrebe parametarskih entiteta u HTML DTD (2)

- Deklaracija lista atributa

```
<!-- Atributi horizont. poravnanja sadržaja ćelije -->
<!ENTITY % cell.halign
  " align (left | center | right) left ">
<!-- Atributi vertikalnog poravnanja sadržaja ćelije -->
<!ENTITY % cell.valign
  " valign (top | middle | bottom) middle">

<!ATTLIST tr      --red tabele--
  %cell.halign;    --horizont. poravnanje u ćeliji--
  %cell.valign;    --vertikal. poravnanje u ćeliji-->
<!ATTLIST td      --ćelija podataka--
  nowrap (nowrap) #IMPLIED
                  --sprečava namotavanje redova --
  rowspan NMTOKEN 1 --broj redova koje pokriva ćelija--
  colspan NMTOKEN 1 --broj kolona koje pokriva ćelija--
  %cell.halign;    --horizont. poravnanje u ćeliji--
  %cell.valign;    --vertikal. poravnanje u ćeliji--
  width %Pixels; #IMPLIED --predložena širina ćelije--
  height %Pixels; #IMPLIED --predložena visina ćelije-->
```

kad god postoji element koji će imati
atribute **align** i **valign**, navode se
parametarski entiteti **cell.halign** i
cell.valign



Opšti entiteti

- Opšti entiteti se deklarišu u okviru DTD-a i koriste prvenstveno u okviru **tela XML dokumenta**
- Mogu se koristiti u okviru DTD-a samo ako će se u krajnjoj instanci uključiti u telo XML dokumenta
 - na primer, kao podrazumevana vrednost atributa
- Oni **ne mogu** da obezbede tekst za sam DTD
- Njihova osnovna namena je da olakšaju pripremu i čitljivost XML dokumenta
- Prema mestu deklaracije, mogu biti:
 - unutrašnji
 - spoljašnji

Unutrašnji entiteti (1)

- **Unutrašnji entiteti** su entiteti kod kojih je tekst zamene niska karaktera zadata kao **literal**
- Njihova deklaracija je oblika:
<!ENTITY *ime_entiteta tekst_entiteta*>
- Tekst entiteta je tekst koji posle parsiranja celog dokumenta zamenjuje referencu entiteta
- Opšti entitet se referiše najčešće u okviru XML dokumenta na sledeći način:
&ime_entiteta;

Unutrašnji entiteti (1)

- **Unutrašnji entiteti** su entiteti kod kojih je tekst zamene niska karaktera zadata kao **literal**
- Njihova deklaracija je oblika:
<!ENTITY ime_entiteta tekst_entiteta>
- Tekst entiteta je tekst koji posle parsiranja celog dokumenta zamenjuje referencu entiteta
- Opšti entitet se referiše najčešće u obliku
&ime_entiteta;

već viđeno
kao karakterski entitet!

na

Unutrašnji entiteti (2)

- Na primer, ako se opšti entitet DDT deklariše u okviru DTD-a
`<!ENTITY DDT "Dichlorodiphenyltrichlorethan">`
onda se on može referisati u tekstu XML dokumenta
DDT, ili &DDT;, je pesticid koji je izazvao veliko zagađenje.
- Rezultat za aplikaciju koja obrađuje dokument (recimo, za pregledač) je:
DDT, ili **Dichlorodiphenyltrichlorethan**, je pesticid koji je izazvao veliko zagađenje.
- Entiteti mogu da sadrže i tekst i XML elemente
- Bitno je da i tekst zamene mora da bude **dobro formiran**, odnosno nije dozvoljeno da početna etiketa nekog elementa bude u tekstu zamene entiteta, a završna izvana njega ili u tekstu zamene nekog drugog entiteta

Your company name

Primer unutrašnjeg entiteta (2)

```
<!ELEMENT urednik (ime)
<!ENTITY urednik_ime "<ime>Milena Jovanović">
```

```
<časopis>
  <urednik>
    <ime>&urednik_ime;</ime>
  </urednik>
</časopis>
```

```
<časopis>
  <urednik>
    <ime><ime>Milena Jovanović</ime>
  </urednik>
</časopis>
```

posle zamene, nije
dobro formiran
dokument!

Primer unutrašnjeg entiteta (3)

```
<!ELEMENT urednik (ime)
```

```
<!ENTITY urednik_ime "<ime>Milena Jovanović</ime>">
```

```
<časopis>
```

```
  <urednik>&urednik_ime;</urednik>
```

```
</časopis>
```

```
<časopis>
```

```
  <urednik>
```

```
    <ime>Milena Jovanović</ime>
```

```
  </urednik>
```

```
</časopis>
```

posle zamene,
dokument je dobro
formiran

Primer unutrašnjeg entiteta (1)

<!ENTITY zaglavlje '

<naziv>Svet kompjutera</naziv>

<glavni_urednik>Zoran Mošorinski</glavni_urednik>

<sporedni_urednik>Tihomir Stančević</sporedni_urednik>

<sporedni_urednik>Dragan Kosovac</sporedni_urednik>

<lektor>Miloš Cvetković</lektor>

<ilustracije>Predrag Milićević</ilustracije>

<opis>Kompjuterski časopis "Svet kompjutera" bavi se temama koje obuhvataju kućne i PC računare

i njihovu upotrebu za posao i zabavu. </opis>

'>

Primer unutrašnjeg entiteta (1)

```
<novine>  
  &zaglavlje;  
  <clanci>  
    <clanak>Tekst prvog clanka...</clanak>  
    <clanak>Tekst drugog clanka...</clanak>  
  </clanci>  
</novine>
```

Your company name



Spoljašnji entiteti (1)

- Oni se deklarišu pomoću ENTITY deklaracije u okviru DTD-a
- Umesto teksta zamene, koristi se ključna reč SYSTEM i URL datoteke u kome je tekst zamene

<!ENTITY ime_entiteta SYSTEM URL_teksta_entiteta>

- Na primer, zaglavlje neke web stranice (npr, imena članova novinske redakcije) može biti smešten u zasebnu datoteku

**<!ENTITY zaglavlje SYSTEM
"entiteti/zaglavlje.xml">**

a može se koristiti i relativni URL

<!ENTITY navigacija SYSTEM "../entiteti/zaglavlje.xml">

Spoljašnji entiteti (2)

- Sadržaj ovako deklarisanog spoljašnjeg entiteta se **parsira**, što znači, da može da se dogodi da sadrži etikete koje će parser na odgovarajući način interpretirati, a rezultat proslediti aplikaciji
- Osim toga, datoteka ne sme da sadrži karaktere "<" i "&", osim u značenju početka etikete, odnosno, reference entiteta
- Dalje, sadržaj datoteke mora da bude potencijalno **dobro formiran** - ne mora da sadrži jedan koreni element, ali svaka otvorena etiketa mora imati svoju zatvorenu etiketu u istoj datoteci

Spoljašnji entiteti (3)

Spoljašnji entitet može da sadrži XML deklaraciju:

- U deklaraciji teksta obavezan je atribut **encoding**, jer se može koristiti druga kodna šema u odnosu na osnovni XML dokument, atribut **version** je opcioni, a atribut **standalone** se ne sme koristiti
- Dozvoljene deklaracije teksta bile bi:

<?xml version="1.0" encoding="ISO-8859-1"?>

ili

<?xml encoding="ISO-8859-1"?>

Primer spoljašnjeg entiteta (1)

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE novine [
  ...
  <!-- Sadržaj zaglavlja je u spoljasmem dokumentu: -->
  <!ENTITY zaglavlje SYSTEM "zaglavlje.xml">
]>
<novine>
  &zaglavlje;
  ...
</novine>
```


zaglavlje.xml

```
<?xml version="1.0" encoding="UTF-8"?>  
<naziv>Svet kompjutera</naziv>  
<glavni_urednik>Zoran Mošorinski</glavni_urednik>  
<sporedni_urednik>Tihomir Stančević</sporedni_urednik>  
<sporedni_urednik>Miodrag Kuzmanović</sporedni_urednik>  
<sporedni_urednik>Dragan Kosovac</sporedni_urednik>  
<lektor>Miloš Cvetković</lektor>  
<ilustracije>Predrag Milićević</ilustracije>  
<opis>Kompjuterski časopis "Svet kompjutera" bavi se temama  
koje obuhvataju ....</opis>
```



Javni identifikatori DTD-a (1)

- Omogućavaju da se identifikuje i koristi neki DTD koji je stavljen u javnu upotrebu
- Javni identifikator se koristi na sledeći način:
 <!
 DOCTYPE *ime_korena* PUBLIC "*Detalji_o_DTD*" "*Opcioni_URL_DTDa*">
- Detalji o javnom DTD-u zadaju se u obliku
 "**prefiks//autor//opis//KodOd2SlovaJezikaKojiSeKoristi**"
- Prefiks može da bude
 - „**ISO DTD**“ je ISO standard
 - „**+**“ DTD je standard neke druge organizacije za standardizaciju
 - „**-**“ DTD nije propisala nijedna organizacija za standardizaciju

Javni identifikatori DTD-a (2)

- Na primer, korišćenje DTD-a za seminarski:
<!DOCTYPE text PUBLIC "-//Branislava Sandrih//Seminarski//EN"
"<http://www.fil.bg.ac.rs/branislava/seminarski/ip4/resursi/vesti.dtd>">
- U XMLCopyEditor-u mora da se odobri pristup internetu: Edit → Preferences → Enable network access for XML validation



Javni identifikatori karakterskih entiteta (1)

Osim javnih DTD-a, postoje i **javne kolekcije entiteta** koje se mogu uključiti u neki DTD

Na primer, kolekcija entiteta koja omogućava da se koriste entiteti za simbole iz gornje kodne strane karakterskog skupa Latin 1 (ISO-8859-1)

```
<!ENTITY % HTMLlat1 PUBLIC  
    "-//W3C//ENTITIES Latin 1 for XHTML//EN"  
    "http://www.w3.org/TR/xhtml1/DTD/xhtml-lat1.ent">
```

upotreba: **%HTMLlat1;**

Your company name

Javni identifikatori karakterskih entiteta (2)

- Sledeća kolekcija omogućava korišćenje matematičkih simbola i grčkih slova:

<!ENTITY % HTMLsymbol PUBLIC

"-//W3C//ENTITIES Symbols for XHTML//EN"

"<http://www.w3.org/TR/xhtml1/DTD/xhtml-symbol.ent>">

upotreba: **%HTMLsymbol;**



Javni identifikatori karakterskih entiteta (3)

- Sledeća kolekcija omogućava korišćenje ćirilice (ovo je ISO standard):

<!ENTITY % Cirlica PUBLIC

"ISO 8879:1986//ENTITIES Non-Russian Cyrillic//EN//XML"

"http://www.oasis-open.org/docbook/xmlcharent/0.3/iso-cyr2.ent">

upotreba: **%Cirlica;**



Uslovno uključivanje i isključivanje (1)

- U XML-u postoji mogućnost za uslovno uključivanje ili isključivanje nekih delova XML dokumenta
- Delovi koji se uslovno uključuju ili isključuju mogu sadržati DTD ili primerak dokumenta
- Na primer, ako se u DTD-u nađe sledeći odeljak:

```
<![IGNORE[  
  <!ELEMENT book (title,author,summary)>  
]]>
```

efekat će biti isti kao da deklaracije elementa book u DTD-u nema
- Ako se u istom odeljku ključna reč IGNORE zameni sa INCLUDE:

```
<![INCLUDE[  
  <!ELEMENT book (title,author,summary)>  
]]>
```

element book biće uključen u DTD



Uslovno uključivanje i isključivanje (2)

- Pogodno je što se element book, i s njim povezani elementi može po potrebi uključivati i isključivati
- Na primer, ako se deklariše parametarski entitet

<!ENTITY % draft "INCLUDE">

onda umesto ključne reči može da se koriste referenca parametarskog entiteta

<![%draft;

<!ELEMENT book (title,author,summary)>

]]>

Uslovno uključivanje i isključivanje (3)

- Ključne reči **INCLUDE** i **IGNORE** mogu se koristiti i kada svi delovi XML dokumenta ne ulaze u sve njegove verzije
- U zavisnosti od toga da li je reč o završenoj knjizi, ili ne:

```
<!ENTITY % draft "INCLUDE">
```

```
<!ENTITY % final "IGNORE">
```

```
<![%draft;
```

```
<!ELEMENT book (title,author,summary)>
```

```
]]>
```



```
<![%final;
```

```
<!ELEMENT book (comments*,title,author,summary)>
```

```
]]>
```