

Exercise 1.6

a) DTD, no attributes.

• element names $E_1 \dots E_n$

↳ complete tree starting symbol: S

$$S \rightarrow E_1 | \dots | E_n$$

• for every element type:

↳ ELEMENT E_i *content model*

• if content model \equiv EMPTY:

$$E_i \rightarrow \langle \text{element}_i \rangle$$

• if content model \equiv (A|P(DATA))

$$E_i \rightarrow \langle \text{element}_i \rangle P \langle \text{element}_i \rangle$$

$$P \rightarrow \epsilon | "a" P | "b" P | \dots | "z" P | \dots$$

Nov 29-10:12

• content model: complex expression β :

$$E_i \rightarrow \langle \text{element}_i \rangle C_\beta \langle \text{element}_i \rangle$$

$$\text{if } \beta = (\beta_1, \beta_2):$$

$$C_\beta \rightarrow C_{\beta_1} C_{\beta_2}$$

$$\text{if } \beta = (\beta_1 | \beta_2):$$

$$C_\beta \rightarrow C_{\beta_1} | C_{\beta_2}$$

$$\text{if } \beta = (\beta_1^*):$$

$$C_\beta \rightarrow \epsilon | C_{\beta_1} C_\beta$$

$$\text{if } \beta = (\beta_1^+):$$

$$C_\beta \rightarrow C_{\beta_1} (\epsilon | C_{\beta_1}^+)$$

$$\text{if } \beta = (\beta_1^?):$$

$$C_\beta \rightarrow \epsilon | C_{\beta_1} \quad (a|b)^*$$

⇒ finite number of *context-free* production rules

⇒ context free grammar → XML: DTD without attrs is context free

Nov 29-10:27

b) attributes: production

- extend opening tag with all permutations of attrs (care for #REQ/#IMPL)
- Values of attr: characters (care for whitespace)
- Context free

+:

- IDs must be unique, (1)
- DDFs must not be dangling (2)
- not context-free

real parsing: basically context free,

→ dictionary with all ids (1)

- parse complete: check uniqueness
- at the end: check (2)

Nov 29-10:29

(1) not context free

proof: there is no context free grammar (in the general case)

→ use the "pumping lemma"

backed:

Context free grammar

$$N \rightarrow \alpha_1 \alpha_2 \dots \alpha_n$$

e.g. $N \rightarrow z_1 z_2 N_1 z_3$

frame: finite number of such rules

large documents require recursion!

Nov 29-10:43

Purpose lemma:
 arbitrary long "word" (=XML doc)

$\langle a \dots a \ b \dots b \rangle \langle / a \dots a \ b \dots b \rangle$

1) vwx only in opening tag (or part in closing tag)
 $\rightarrow u v^{+++} w x^{+++} y$
 $\equiv \langle a_{\leq n} a^{++} \underbrace{qb b^{+++}}_{\text{much longer opening tag}} \cdot b \rangle \langle a^n b^n \rangle$

2) vwx "in the middle", v in opening tag, x in closing tag
 $u v^{+++} w x^{+++} y$
 $\langle a^n b \cdot b b^{+++} b s \rangle \langle / a s a^{+++} a s b^n \rangle$

3) v (or x) duplicate something of ...
 $u v^{+++} w x^{+++} y$
 $\rightarrow v = "bs" \langle / as "$
 $\langle a^n b s \ \underbrace{bs}_{v} \rangle \langle / as \ \underbrace{bs}_{v} \rangle \langle / as \ \underbrace{as}_{x} b s \rangle$

Nov 29-10:46



Nov 29-10:51