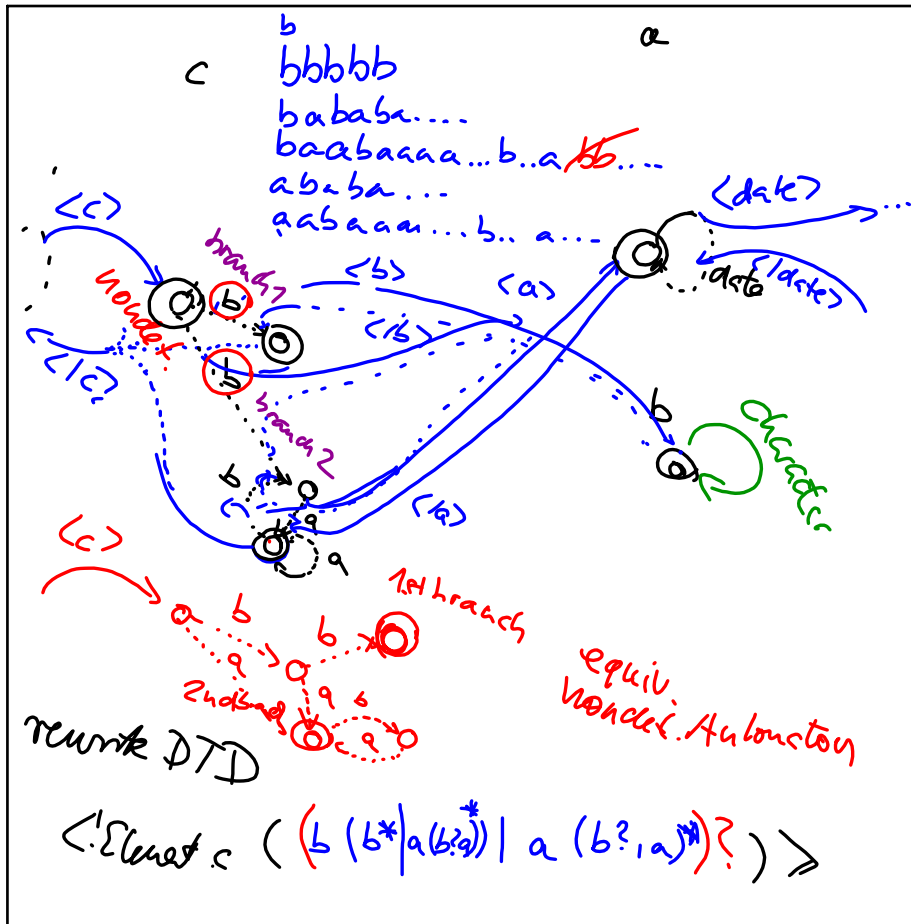


Dez 3-10:07



Dez 3-10:42

for each element name:

basic idea:

$\langle \text{elementname}_i \rangle \dots \langle \text{elementname}_i \rangle$

hinten anfügen ✓

formally:

$\Sigma = \{ \langle, \rangle, |, =, a, \dots, z, A, Z, 0, \dots, 9, \dots \}$ ← Terminal symbols

$\cup \{ \text{elementname}_1, \text{elementname}_2 \} \cup \dots$

Nonterminal symbols

$\{ S, E_1, \dots, E_n, C_1, \dots, C_n, F_1, \dots, P \}$

Productions // $S \rightarrow$ the root element

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

$E_1 \rightarrow \langle \text{elementname}_1 \rangle C_1$ ← $\langle \text{elementname}_1 \rangle$ some regexp.

$E_2 \rightarrow \langle \text{elementname}_2 \rangle$

\vdots

$C_1 \rightarrow E_1 E_2$

$C_3 \rightarrow \dots$

$C_4 \rightarrow \dots$

ex: $\langle !ELEMENT E_1 (E_2 E_3) \rangle$

depending on the DTD

Dez 3-11:10

for each C_i, C_j, \dots depends on the DTD

if $\beta = \#PCDATA$ $C_\beta \rightarrow P$

$P \rightarrow P_1 | P_2 | \dots | P_n | A, \dots, z$

$aP | bP | \dots | zP | \dots | \epsilon$

if $\beta = (B_1 B_2)$ $C_\beta \rightarrow C_{B_1} C_{B_2}$

if $\beta = (B_1 | B_2)$ $C_\beta \rightarrow C_{B_1} | C_{B_2}$

if $\beta = (B^*)$ $C_\beta \rightarrow C_{B_1} C_\beta | \epsilon$

if $\beta = (B_1^+)$ $C_\beta \rightarrow C_{B_1} C_\beta | \epsilon$

if $\beta = (B_1?)$ $C_\beta \rightarrow C_{B_1} (C_\beta | \epsilon)$

if $\beta = E_i$ $C_\beta \rightarrow C_{E_i} | \epsilon$

if $\beta = ANY$ $C_{Any} \rightarrow E_1 C_\beta | E_2 C_\beta | \dots | E_n C_\beta | \epsilon$

in case somewhere E_{ij} should include $ANY + PCDATA$! No PCDATA!

$\langle !ELEMENT E_j (\#PCDATA | \text{element}_1 | \dots | \text{element}_n) \rangle$

(S/Id 157) $\beta = \text{mixed}$

if $\beta = \dots C_{mixed} \rightarrow C_{Any} | C_{mixed} P$

\Rightarrow context free grammar

Dez 3-11:20

attributes?

consider that

$\langle !ATTLIST \ e_1 \ a_1 \ a_2 \ a_3 \ \rangle$

take good as before

a_1 CONST #Required
 a_2 CONST #Required
 a_3 CONST #Implied

$\rightarrow \langle e_1 \ a_1="P" \ a_2="P" \ a_3="P" \ \rangle \langle /e_1 \ \rangle$

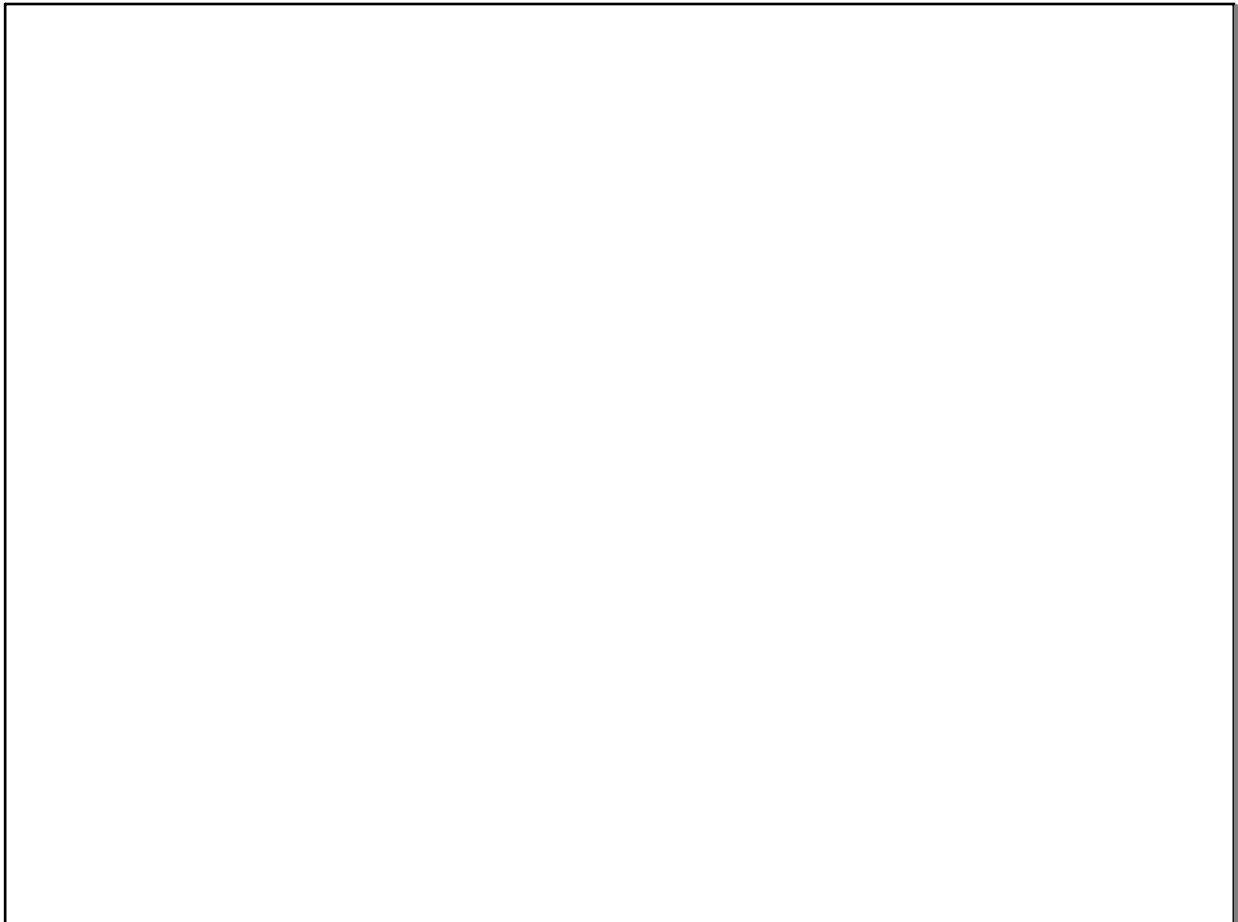
$a_1 a_2$, $a_2 a_1$, $a_2 a_1 a_3$
 $a_3 a_1 a_2$, $a_3 a_2 a_1$
 $a_2 a_3 a_1$, ... ? all permutations
 (with/without opt.)

\Rightarrow still context-free

for ID: at the end of parsing: must be unique
 \Rightarrow needs a dictionary

for REF: after parsing: check with Dict.

Dez 3-11:29



Dez 3-11:42