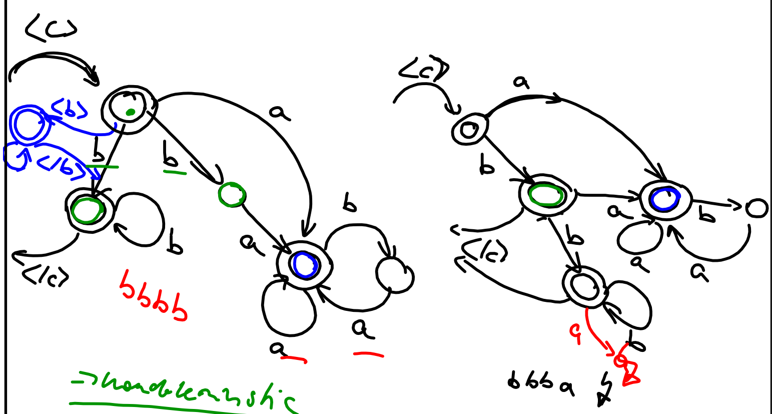


→ nondeterministic

ab.... DFA

⊊ regular language

⊊ context-free lg

↝ push-down automaton

or $\boxed{\text{DFA + stack}}$

Mai 25-10:15

lg is context-free

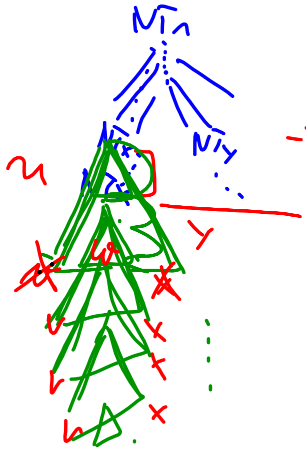
• if there is a (finite) context-free grammar

$NT_1 \rightarrow \, ..NT_x$

$NT_2 \rightarrow \, ...NT_y NT_2$ unprobable recursive

$NT_3 \rightarrow \, ...NT_1$

↝ trees can be arbitrarily deep



→ there must be any recursion

Mai 25-10:45

a) DTD : finite number of element decls.

<! Element elname$_1$> (β) >
<! Element elname$_2$> (β_2) >
⋮

content model

$\beta_1 \equiv elname_2{}^*$

$E_1 \rightarrow \langle elname_1 \rangle\ E2S \langle / \rangle$

$E_2S \rightarrow \varepsilon \mid E_2 E_2S$

$E2 \rightarrow \langle elname_2 \rangle \ldots\ldots \langle / \rangle$

Mai 25-10:50

for each
<! Element elname$_k$ (β) >
add rule

$E_K \rightarrow \langle elname_k \rangle\ \underline{C_\beta}\ \langle / elname_k \rangle$

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

$P \rightarrow$ "a"P | "b"P | … | "z"P | "A"P | … "Z"P | 0P …
not "<"P

if $\beta = \beta_1, \beta_2$: $C_\beta \rightarrow C_{\beta_1} C_{\beta_2}$

if $\beta = \beta_1 | \beta_2$: $C_\beta \rightarrow C_{\beta_1}$, $C_\beta \rightarrow C_{\beta_2}$

if $\beta = \beta_1{}^*$: $C_\beta \rightarrow \varepsilon \mid C_{\beta_1} C_\beta$

if $\beta = \beta_1{}^+$: $C_\beta \rightarrow C_{\beta_1}(C_\beta \mid \varepsilon)$

if $\beta = \beta_1 ?$: $C_\beta \rightarrow \varepsilon \mid C_{\beta_1}$

if $\beta = elname_j$: $C_\beta \rightarrow E_j$

→ context-free grammar for $L_{(a)}$
as long as there is no recursion in the DTD,
it is even regular
→ it's even finite → 'trivial'
the Standard-XML lg. is regular

Mai 25-10:55

other DTD for mondial

<sea><name>North Sea</>
<river><name>Elbe</> </>
<river><name>Weser</>
<river><name>Aller</>
<river><name>Leine</></> </>
..... might be arbitrarily deep

⇒ recursive DTD

<!Element river (name, river*))

Mai 25-11:02

b) Attributes?

<elname$_k$ $\underbrace{a_n= \;\; a_3= \;\; a_1= \ldots}_{\text{not ordered}}$> β </>

<!ATTLIST elname$_k$
fixed
attname$_1$ ....
⋮
attname$_n$ ... >

not ordered

⇒ any permutation of
attname$_1$='P'
⋮
attname$_n$='P'

⇒ E_k → <elname$_1$ attname$_1$='P' ...attname$_n$='P'> β </>
| all permutations → finite
elname$_1$ attname$_n$='P' .... attname$_1$='P'> β </>

Mai 25-11:11

c) XML files contextfree?

$$L_{xml} = \{ \langle \underbrace{\omega_1 \ldots \omega_k}_{\text{char}} \rangle \underbrace{\alpha_1 \ldots \alpha_m}_{\text{something context}} \langle / \underbrace{\omega_1 \ldots \omega_k}_{\text{char}} \rangle \mid \ldots \}$$

Pumping lemma:

some n

following word $z = \langle \underbrace{a \ldots a}_{n} \underbrace{b \ldots b}_{n} \rangle \langle / a \ldots a b \ldots b \rangle$

$\forall\, u \underline{vwx} y = z$

either $|vx| = 0$

or $|vwx| > n$

or $\boxed{\exists i \;\; u v^i w x^i y \notin L_{XML}}$

Mai 25-11:19

choose some n "quite large"

$\langle \underbrace{a \ldots a}_{n} \underbrace{b \ldots b}_{n} \rangle \langle / \underbrace{a \ldots a}_{n} \underbrace{b \ldots b}_{n} \rangle = u \underline{v} w \underline{x} y$

(1) (2)(3)

Assume $|vx| > 1$ and $|vwx| < n$

Then

1. vwx in the opening tag $\Rightarrow$ $vwx \sim aaabbb$

$\rightsquigarrow u v^2 w x^2 y = \langle aaa\,aabbbaaabbb\,bb \ldots \rangle \langle / a \ldots a b \ldots b \rangle$ ↯

2. "$\rangle\langle/$" in v or in x $\to$ duplicate it $\quad v = bb\rangle\langle/aa$

uv^2wx^2z

$\langle a \ldots a b \ldots b \overbrace{bb\rangle\langle/aabb\rangle\langle/aa}^{v^2} \underset{w}{aa} \underset{x}{(ab)^*} \underset{y}{(a..)bbb\rangle}$ ↯

3. $\rangle\langle/$ in $w \to v = bbb$ (from opening tag) $x = aaa$ (from closing tag)

$u v^2 w x^2 y = \langle aa..bb\,bbbbbb\,bb\rangle\langle/aaa\,aaaaa\,abbbb\rangle$ ↯

"typical" pumping lemma proof "in reality"

Mai 25-11:26