

Entrega No1 Analizador sintáctico Lenguaje SXML

Jefferson David Castañeda Carreño

Diego Andrés Borrero Muñoz

1. El analizador deberá trabajar sobre un vocabulario $v = \{<, >, /, ", =\}$, por lo que un carácter distinto deberá ser un error.
2. Tipos de elementos en SXML
 - a. Del tipo `<ID/>` (marca sencilla) y debe haber exactamente 1 en el SXML si son los únicos que aparecen, y no hay de otro tipo.
 - b. Del tipo `<ID {attr}>` (marca de apertura)
 - c. Del tipo `</ID>` (marca de cierre)
 - d. Las marcas de apertura y cierre conforman a su vez un tipo denominado marcas dobles.
3. A partir de una gramática tipo dos, o una BNF debemos verificar que la gramática cumpla la condición:
 - a. la gramática Analizador debe ser LL(1)
 - i. No ser recursiva por la izquierda
 - ii. No ser ambigua
 - iii. Los conjuntos de selección entre producciones deben ser disjuntos.

Partiendo de esta base, generamos una gramática tipo 2:

- D : No terminal desde el que se puede obtener una marca doble
 - A : No terminal desde el que se puede obtener una marca de apertura
 - C : No terminal desde el que se puede obtener una marca de cierre
- S : No terminal desde el que se puede obtener una marca sencilla
- T : No terminal desde el que se puede obtener atributos
- L : No terminal que puede generar cualquier marca
- P : No terminal sobre el que se trabajaran las producciones principales

El conjunto de terminales sobre el que trabajara la gramática es:

- $\{<, >, /, ", =, id\}$ y para la gramática extendida $\{\$ \}$

Cabe aclarar que para una marca de apertura, debe generarse una marca de cierre, por lo que estas dos producciones por separado son invalidas, o si se incluyen pueden resultar en una ambigüedad, para ello preferimos incluir las producciones A y C en D así:

$$\begin{aligned}
P &\rightarrow D \text{ EOF} \\
P &\rightarrow S \text{ EOF} \\
D &\rightarrow \text{ANGI ID } T \text{ ANG D } L \text{ ANGI SLASH ID ANG D} \\
T &\rightarrow \text{ID EQ STRING } T \\
S &\rightarrow \text{ANGI ID } T \text{ SLASH ANG D} \\
L &\rightarrow DL \\
L &\rightarrow SL \\
L &\rightarrow \lambda
\end{aligned}$$

4. Tratamos de reducir las producciones quitando las inútiles, como por ejemplo S, que nos permite llegar a la expresión $\langle id \ T \ / \rangle$, mediante L podemos obtenerla.

AL :

$$\begin{aligned}
P &\rightarrow \text{ANGI ID } T \text{ ANG D } L \text{ ANGI SLASH ID ANG D EOF} \\
P &\rightarrow \text{ANGI ID } T \text{ SLASH ANG D EOF} \\
L &\rightarrow \text{ANGI ID } T \text{ ANG D } L \text{ ANGI SLASH ID ANG D } L \\
T &\rightarrow \text{ID EQ STRING } T \\
L &\rightarrow \text{ANGI ID } T \text{ SLASH ANG D } L \\
L &\rightarrow \lambda
\end{aligned}$$

- Expresamos la gramática AL obtenida como BNF.

$$\begin{aligned}
P &::= [\text{ANGI ID } T \text{ ANG D } L \text{ ANGI SLASH ID ANG D}] \mid [\text{ANGI ID } T \text{ SLASH ANG D}] \\
T &::= \text{ID EQ STRING } T \mid \lambda \\
L &::= \text{ANGI ID } T \text{ ANG D } L \text{ ANGI SLASH ID ANG D } L \mid \text{ANGI ID } T \text{ SLASH ANG D } L \mid \lambda
\end{aligned}$$

- Factorizamos las expresiones

$$\begin{aligned}
P &::= \text{ANGI ID } T (\text{ANG D } L \text{ ANGI SLASH ID } \mid \text{SLASH }) \text{ ANG D} \\
L &::= [\text{ANGI ID } T (\text{ANG D } L \text{ ANGI SLASH ID } \mid \text{SLASH }) \text{ ANG D } L]
\end{aligned}$$

- Aplicamos arden sobre la producción T y L

$$\begin{aligned}
T &::= \{ \text{ID EQ STRING} \} \\
L &::= \{ \text{ANGI ID } T (\text{ANG D } L \text{ ANGI SLASH ID } \mid \text{SLASH }) \text{ ANG D} \}
\end{aligned}$$

- Aplicamos regla de arden simétrica sobre la producción L y la reemplazamos en P

- La nueva gramática es:

$P ::= \text{ANGI ID } T (\text{ANGD } L \text{ SLASH ID } | \text{ SLASH }) \text{ ANGD}$

$T ::= \{ \text{ID EQ STRING} \}$

$L ::= \text{ANGI } \{ \text{ID } T (\text{ANGD } L \text{ SLASH ID } | \text{ SLASH}) \text{ ANGD} \}$