Theory of Automata

CFG Derivations and Parsing

Dr. Sabina Akhtar

Revision

Language of a Grammar

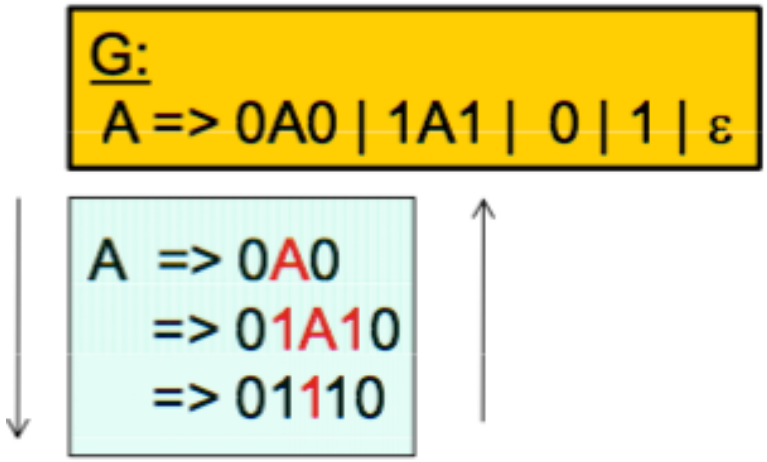
- The language of a CFG, G=(V,T,P,S), denoted by L(G), is the set of terminal strings that have a derivation from the start variable S.
  - $L(G) = \{ w \text{ in } T^* \mid S \Rightarrow^*_G w \}$

□

String membership

- How to say if a string belong to the language defined by a CFG?
  - Derivation

- Example
  - w = 01110
  - Is w a palindrome?



Generalization of derivation

- Derivation is *head* ==> *body*

- A==>X (A derives X in a single step)
- A ==>*$_G$ X (A derives X in a multiple steps)

- <u>Transitivity:</u>

  IFA ==>*$_G$B, and B ==>*$_G$C, THEN A ==>*$_G$ C

The Process

Remaining Input:

(())()

Next
symbol

Steps of leftmost derivation:

B

B -> (RB | ϵ

R ->) | (RR

The Process

Remaining Input:

())()

Next
symbol

Steps of leftmost derivation:

B ==> (RB

B -> (RB | ϵ

R ->) | (RR

The Process

Remaining Input:

))()

Next
symbol

Steps of leftmost derivation:

B ==> (RB

==> ((RRB

B -> (RB | ϵ

R ->) | (RR

The Process

Remaining Input:

)()

Next
symbol

Steps of leftmost derivation:

B ==> (RB

==> ((RRB

==> (()RB

B -> (RB | ϵ

R ->) | (RR

The Process

Remaining Input:

()

Next
symbol

Steps of leftmost derivation:

B ==> (RB

==> ((RRB

==> (()RB

==> (())B

B -> (RB | ϵ

R ->) | (RR

The Process

Remaining Input:

)

Next
symbol

Steps of leftmost derivation:

B ==> (RB

==> ((RRB

==> (()RB

==> (())B

==> (())(RB

B -> (RB | ϵ

R ->) | (RR

The Process

Remaining Input:

Next
symbol

Steps of leftmost derivation:

B ==> (RB

==> ((RRB

==> (()RB

==> (())B

==> (())(RB

==> (())()B

B -> (RB | ϵ R ->) | (RR

The Process

Remaining Input:

Next
symbol

Steps of leftmost derivation:

B ==> (RB

==> ((RRB

==> (()RB

==> (())B

==> (())(RB

==> (())()B

==> (())()

B -> (RB | ϵ

R ->) | (RR

Class Activity

- Provide the leftmost derivation for a * (a + b00) starting from *E* using the grammar below

$$
\begin{array}{lrcl}
1. & E & \rightarrow & I \\
2. & E & \rightarrow & E + E \\
3. & E & \rightarrow & E * E \\
4. & E & \rightarrow & (E) \\
 & & & \\
5. & I & \rightarrow & a \\
6. & I & \rightarrow & b \\
7. & I & \rightarrow & Ia \\
8. & I & \rightarrow & Ib \\
9. & I & \rightarrow & I0 \\
10. & I & \rightarrow & I1
\end{array}
$$

Solution

$$E \Rightarrow E * E \Rightarrow I * E \Rightarrow a * E \Rightarrow$$

$$a * (E) \Rightarrow a * (E + E) \Rightarrow a * (I + E) \Rightarrow a * (a + E) \Rightarrow$$

$$a * (a + I) \Rightarrow a * (a + I0) \Rightarrow a * (a + I00) \Rightarrow a * (a + b00)$$

Class Activity

- Provide the leftmost derivation for

(a101 + b1) * (a1 + b)

1. $E \rightarrow I$
2. $E \rightarrow E + E$
3. $E \rightarrow E * E$
4. $E \rightarrow (E)$
5. $I \rightarrow a$
6. $I \rightarrow b$
7. $I \rightarrow Ia$
8. $I \rightarrow Ib$
9. $I \rightarrow I0$
10. $I \rightarrow I1$

Derivation Types

- Left-most derivation
- Right-most derivation

Left-most & Right-most Derivation Styles

G:
E => E+E | E*E | (E) | F
F => aF | bF | 0F | 1F | ε

Derive the string a*(ab+10) from G:

$E =^{*}\Rightarrow_{G} a*(ab+10)$

Left-most derivation:

Always substitute leftmost variable

- E
- ==> E * E
- ==> F * E
- ==> aF * E
- ==> a * E
- ==> a * (E)
- ==> a * (E + E)
- ==> a * (F + E)
- ==> a * (aF + E)
- ==> a * (abF + E)
- ==> a * (ab + E)
- ==> a * (ab + F)
- ==> a * (ab + 1F)
- ==> a * (ab + 10F)
- ==> a * (ab + 10)

Right-most derivation:

Always substitute rightmost variable

- E
- ==> E * E
- ==> E * (E)
- ==> E * (E + E)
- ==> E * (E + F)
- ==> E * (E + 1F)
- ==> E * (E + 10F)
- ==> E * (E + 10)
- ==> E * (F + 10)
- ==> E * (aF + 10)
- ==> E * (abF + 0)
- ==> E * (ab + 10)
- ==> F * (ab + 10)
- ==> aF * (ab + 10)
- ==> a * (ab + 10)

Class Activity

- Provide the right-most derivation for

  (a101 + b1) * (a1 + b)

1. $E \rightarrow I$
2. $E \rightarrow E + E$
3. $E \rightarrow E * E$
4. $E \rightarrow (E)$

5. $I \rightarrow a$
6. $I \rightarrow b$
7. $I \rightarrow Ia$
8. $I \rightarrow Ib$
9. $I \rightarrow I0$
10. $I \rightarrow I1$

Exercise

[A]**2.3** Answer each part for the following context-free grammar G.

$$
\begin{aligned}
R &\to XRX \mid S \\
S &\to \mathtt{a}T\mathtt{b} \mid \mathtt{b}T\mathtt{a} \\
T &\to XTX \mid X \mid \varepsilon \\
X &\to \mathtt{a} \mid \mathtt{b}
\end{aligned}
$$

a. What are the variables of G?

b. What are the terminals of G?

c. Which is the start variable of G?

d. Give three strings in $L(G)$.

e. Give three strings *not* in $L(G)$.

f. True or False: $T \Rightarrow \mathtt{aba}$.

g. True or False: $T \stackrel{*}{\Rightarrow} \mathtt{aba}$.

h. True or False: $T \Rightarrow T$.

i. True or False: $T \stackrel{*}{\Rightarrow} T$.

j. True or False: $XXX \stackrel{*}{\Rightarrow} \mathtt{aba}$.

k. True or False: $X \stackrel{*}{\Rightarrow} \mathtt{aba}$.

l. True or False: $T \stackrel{*}{\Rightarrow} XX$.

m. True or False: $T \stackrel{*}{\Rightarrow} XXX$.

n. True or False: $S \stackrel{*}{\Rightarrow} \varepsilon$.

o. Give a description in English of $L(G)$.

Exercise

2.4 Give context-free grammars that generate the following languages. In all parts, the alphabet Σ is $\{0,1\}$.

[A] **a.** $\{w \mid w$ contains at least three 1s$\}$

b. $\{w \mid w$ starts and ends with the same symbol$\}$

c. $\{w \mid$ the length of w is odd$\}$

[A] **d.** $\{w \mid$ the length of w is odd and its middle symbol is a 0$\}$

e. $\{w \mid w = w^{\mathcal{R}}$, that is, w is a palindrome$\}$

f. The empty set

References

- Book Chapter
- Lectures from Stanford University
  - http://infolab.stanford.edu/~ullman/ialc/spr10/spr10.html#LECTURE%20NOTES
- Lectures from Washington State University
  - http://www.eecs.wsu.edu/~ananth/CptS317/Lectures/