CSCI 301 - Lecture 24: CFG Application: Parsing

Given a grammar G and a string w, is $w \in L(G)$?

Of practical interest if you ever want to:

- write and read data to a file or network
- write an interpreter or compiler for a programming language
- extract meaning from structured data

This is called **parsing**, and it's (in general) **hard**!

Why? There can be **many ways** to derive the same string.

Example: consider a grammar for arithmetic expressions.

$V = \{E\}$

$\Sigma = \{0, 1, \ldots, 9, +, -, *, /, (,)\}$

$S = E$

$R =$

$E \to E + E$
$\mid E - E$
$\mid E * E$
$\mid E / E$
$\mid (E)$
$\mid 0 \mid 1 \mid 2 \mid \ldots \mid 9$

$L(G)$ includes:

$1 + 1$

$(4 + 7) * (6 / 2)$

$L(G)$ does not include:

$1 +$

$2 * 4)$

$E \to E + E$
$\quad | \; E - E$
$\quad | \; E * E$
$\quad | \; E / E$
$\quad | \; (E)$
$\quad | \; 0 | 1 | 2 | \cdots | 9$

String: $w = 1 + 1 * 4$

Derivations:

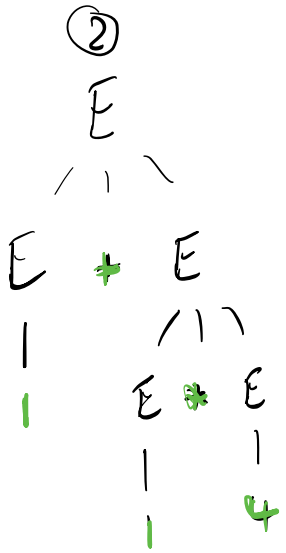
① $E \Rightarrow E + E$
$E + E * E$
$1 + E * E$
$1 + 1 * E$
$1 + 1 * 4$

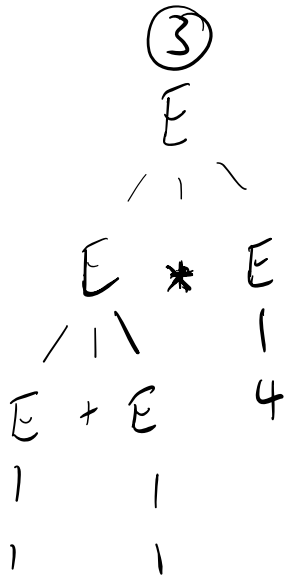
② leftmost
$E \Rightarrow E + E$
$1 + E$
$1 + E * E$
$1 + 1 * E$
$1 + 1 * 4$

③ rightmost
$E \Rightarrow E * E$
$E * 4$
$E + E * 4$
$E + 1 * 4$
$1 + 1 * 4$

Parse trees:

①


②


③


Definitions

left-most derivation : each step applies a rule to the leftmost nonterminal.

right-most derivation : each step applies a rule to the rightmost nonterminal.

Def: A grammar is ambiguous if some string w has more than one parse tree.

Equivalently: A grammar is ambiguous if there is more than one left-most derivation for some string.

Ex. PtA

Parsing Techniques - Overview

Given a grammar $G = (V, \Sigma, R, S)$ and a string $w \in \Sigma^*$, determine whether $w \in L(G)$, i.e., whether $S \Rightarrow^* w$.

Two broad categories of approaches:

- Top-down: start with S, apply rules until you've derived w.
- Bottom-up: start with w, apply substitutions "backwards" to get S.

Our focus

In both cases, the crux is: which rule should we apply?

Which nonterminal? leftmost or rightmost

Which rule? ??

General, brute force solution: backtracking

$E \to E + E$
$\quad | \; E - E$
$\quad | \; E * E$
$\quad | \; E / E$
$\quad | \; (E)$
$\quad | \; 0 | 1 | 2 | \ldots | 9$

$E \Rightarrow E + E$

$E - E + E$ ✗

$1 * 1 + 4$

Can we do better?

Top-down:

- LL(k)
  - left-to-right
  - leftmost
  - lookahead
- LL(1)
  - Lab 8

Bottom-up:

- LR(k)
  - rightmost
- LALR

Left Recursion

$w = aabb$

$$S \to aS \mid bS \mid \varepsilon$$

$$S \to Sa \mid Sb \mid \varepsilon$$