UMB CS 622

PDA ⇔ CFL

Monday, March 25, 2024
Announcements

• HW 5 in
  • Due Mon 3/25 12pm noon

• HW 6 out
  • Due Mon 4/1 12pm noon

(AN UNMATCHED LEFT PARENTHESIS CREATES AN UNRESOLVED TENSION THAT WILL STAY WITH YOU ALL DAY.)
## Regular Language vs CFL Comparison

<table>
<thead>
<tr>
<th>Regular Languages</th>
<th>Context-Free Languages (CFLs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regular Expression</td>
<td>Context-Free Grammar (CFG)</td>
</tr>
<tr>
<td>describes a Regular Lang</td>
<td>describes a CFL</td>
</tr>
</tbody>
</table>
## Regular Language vs CFL Comparison

<table>
<thead>
<tr>
<th>Regular Languages</th>
<th>Context-Free Languages (CFLs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regular Expression</td>
<td>Context-Free Grammar (CFG)</td>
</tr>
<tr>
<td><code>describes</code> a Regular Lang</td>
<td><code>describes</code> a CFL</td>
</tr>
<tr>
<td>Finite State Automaton (FSM)</td>
<td>???</td>
</tr>
<tr>
<td><code>recognizes</code> a Regular Lang</td>
<td><code>recognizes</code> a CFL</td>
</tr>
</tbody>
</table>
# Regular Language vs CFL Comparison

<table>
<thead>
<tr>
<th>Regular Languages</th>
<th>Context-Free Languages (CFLs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regular Expression</td>
<td>Context-Free Grammar (CFG)</td>
</tr>
<tr>
<td>describes a Regular Lang</td>
<td>describes a CFL</td>
</tr>
<tr>
<td>Finite State Automaton (FSM)</td>
<td>Push-down Automata (PDA)</td>
</tr>
<tr>
<td>recognizes a Regular Lang</td>
<td>recognizes a CFL</td>
</tr>
</tbody>
</table>

**Definitions:**
- **Regular Expression**
  - describes a Regular Lang
- **Finite State Automaton (FSM)**
  - recognizes a Regular Lang
- **Context-Free Grammar (CFG)**
  - describes a CFL
- **Push-down Automata (PDA)**
  - recognizes a CFL
# Regular Language vs CFL Comparison

<table>
<thead>
<tr>
<th>Regular Languages</th>
<th>Context-Free Languages (CFLs)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Regular Expression</strong></td>
<td><strong>Context-Free Grammar (CFG)</strong></td>
</tr>
<tr>
<td>describes a Regular Lang</td>
<td>describes a CFL</td>
</tr>
<tr>
<td><strong>Finite State Automaton (FSM)</strong></td>
<td><strong>Push-down Automata (PDA)</strong></td>
</tr>
<tr>
<td>recognizes a Regular Lang</td>
<td>recognizes a CFL</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Proved:</th>
<th>Must Prove:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regular Lang $\Leftrightarrow$ Regular Expr</td>
<td>CFL $\Leftrightarrow$ PDA  ???</td>
</tr>
</tbody>
</table>
A lang is a CFL iff some PDA recognizes it

⇒ If a language is a CFL, then a PDA recognizes it
  • We know: A CFL has a CFG describing it (definition of CFL)
  • To prove this part, show: the CFG has an equivalent PDA

⇐ If a PDA recognizes a language, then it’s a CFL
Shorthand: Multi-Symbol Read Transition
Shorthand: Multi-Stack Push Transition

Note the reverse order of pushes
**CFG→PDA (sketch)**

- Construct PDA from CFG such that:
  - PDA accepts input only if CFG generates it
- PDA:
  - simulates generating a string with CFG rules
  - by (nondeterministically) **trying all rules** to find the right ones

![Diagram of PDA with states q_start, q_loop, q_accept and transitions for starting, epsilon, epsilon→S$, epsilon, A→w for rule A→w, a, a→ε for terminal a, and epsilon, $→ε.
CFG\rightarrow PDA (sketch)

- Construct PDA from CFG such that:
  - PDA accepts input only if CFG generates it
- PDA:
  - simulates generating a string with CFG rules
  - by (nondeterministically) trying all rules to find the right ones

Convert: every CFG rule to PDA “loop” transition(s) that:
- Pops LHS variable
- Pushes RHS

\[ \varepsilon, A \rightarrow w \quad \text{for rule } A \rightarrow w \]
\[ a, a \rightarrow \varepsilon \quad \text{for terminal } a \]

Convert: every terminal to “loop” transition that:
- Reads input char
- Pops matching char on stack
**CFG→PDA** (sketch)

- Construct PDA from CFG such that:
  - PDA accepts input only if CFG generates it
- PDA:
  - simulates generating a string with CFG rules
  - by (nondeterministically) trying all rules to find the right ones

---

**Diagram:**
- $q_{start}$
- $q_{loop}$
- $q_{accept}$

- $\epsilon, \epsilon \rightarrow S$  
  - **push start variable onto stack**
- $\epsilon, A \rightarrow w$  
  - **If: stack top is variable** $A$, **pop** and ...
- $\epsilon, \epsilon \rightarrow \epsilon$  
  - **... push rule's right-sides** (nondeterministically)
- $a, a \rightarrow \epsilon$  
  - **If: stack top is terminal** $a$, **pop** and ...
- $\epsilon, \epsilon \rightarrow \epsilon$  
  - **... read matching input**
Example **CFG→PDA**

- **Push start variable onto stack**
- If: stack top is variable $S$, pop $S$ and...
- ...push rule right-sides (in rev order)

**Grammar Rules**:
- $S \rightarrow aTb | b$
- $T \rightarrow Ta | \varepsilon$

**PDA States**:
- $q_{start}$
- $q_{loop}$
- $q_{accept}$
Example **CFG → PDA**

\[ S \rightarrow aTb | b \]
\[ T \rightarrow Ta | \varepsilon \]
Example **CFG→PDA**

$$S \rightarrow aTb \mid b$$

$$T \rightarrow Ta \mid \varepsilon$$

If: stack top is terminal, **pop** and read matching input
Example **CFG→PDA**

**CFG Rules:**
- $S \rightarrow aTb \mid b$
- $T \rightarrow Ta \mid \varepsilon$

**PDA Example:**

<table>
<thead>
<tr>
<th>State</th>
<th>Input</th>
<th>Stack</th>
<th>Equiv Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>$q_{start}$</td>
<td>$aab$</td>
<td>$\varepsilon$</td>
<td>$S \rightarrow aTb$</td>
</tr>
<tr>
<td>$q_{loop}$</td>
<td>$aab$</td>
<td>$Sb$</td>
<td>$S \rightarrow aTb$</td>
</tr>
<tr>
<td>$q_{loop}$</td>
<td>$aab$</td>
<td>$TaS$</td>
<td>$T \rightarrow Ta$</td>
</tr>
<tr>
<td>$q_{loop}$</td>
<td>$ab$</td>
<td>$TbS$</td>
<td>$T \rightarrow Ta$</td>
</tr>
<tr>
<td>$q_{loop}$</td>
<td>$ab$</td>
<td>$abS$</td>
<td>$T \rightarrow \varepsilon$</td>
</tr>
<tr>
<td>$q_{loop}$</td>
<td>$b$</td>
<td>$bS$</td>
<td>$T \rightarrow \varepsilon$</td>
</tr>
<tr>
<td>$q_{accept}$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Example Derivation using CFG:

1. $S \rightarrow aTb$  (using rule $S \rightarrow aTb$)
2. $\Rightarrow aTab$ (using rule $T \rightarrow Ta$)
3. $\Rightarrow aab$ (using rule $T \rightarrow \epsilon$)

If: stack top is variable $S$, pop $S$ and push rule right-sides (in rev order)

PDA Example:

<table>
<thead>
<tr>
<th>State</th>
<th>Input</th>
<th>Stack</th>
<th>Equiv Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>$q_{start}$</td>
<td>aab</td>
<td>$\epsilon$</td>
<td>$S \rightarrow aTb$</td>
</tr>
<tr>
<td>$q_{loop}$</td>
<td>aab</td>
<td>$S$</td>
<td>$S \rightarrow aTb$</td>
</tr>
<tr>
<td>$q_{loop}$</td>
<td>aab</td>
<td>$aTb$</td>
<td>$S \rightarrow aTb$</td>
</tr>
<tr>
<td>$q_{loop}$</td>
<td>ab</td>
<td>$Tb$</td>
<td>$T \rightarrow Ta$</td>
</tr>
<tr>
<td>$q_{loop}$</td>
<td>ab</td>
<td>$ab$</td>
<td>$T \rightarrow \epsilon$</td>
</tr>
<tr>
<td>$q_{loop}$</td>
<td>b</td>
<td>$b$</td>
<td>$T \rightarrow \epsilon$</td>
</tr>
<tr>
<td>$q_{accept}$</td>
<td></td>
<td>$$</td>
<td>$T \rightarrow \epsilon$</td>
</tr>
</tbody>
</table>
Example Derivation using CFG:
\[ S \Rightarrow aTb \] (using rule \( S \Rightarrow aTb \))
\[ \Rightarrow aTab \] (using rule \( T \Rightarrow Ta \))
\[ \Rightarrow aab \] (using rule \( T \Rightarrow \varepsilon \))

PDA Example

<table>
<thead>
<tr>
<th>State</th>
<th>Input</th>
<th>Stack</th>
<th>Equiv Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>( q_{start} )</td>
<td>aab</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( q_{loop} )</td>
<td>aab</td>
<td>$S$</td>
<td></td>
</tr>
<tr>
<td>( q_{loop} )</td>
<td>aab</td>
<td>$a\gamma b$</td>
<td>( S \Rightarrow aTb )</td>
</tr>
<tr>
<td>( q_{loop} )</td>
<td>ab</td>
<td>$7b$</td>
<td></td>
</tr>
<tr>
<td>( q_{loop} )</td>
<td>ab</td>
<td>$ab\gamma$</td>
<td>( T \Rightarrow Ta )</td>
</tr>
<tr>
<td>( q_{loop} )</td>
<td>ab</td>
<td>$ab\gamma$</td>
<td>( T \Rightarrow \varepsilon )</td>
</tr>
<tr>
<td>( q_{loop} )</td>
<td>b</td>
<td>$b\gamma$</td>
<td></td>
</tr>
<tr>
<td>( q_{accept} )</td>
<td></td>
<td>$\gamma$</td>
<td></td>
</tr>
</tbody>
</table>

If stack top is terminal, pop and read matching input.
Example **CFG→PDA**

**Example Derivation using CFG:**

- $S \Rightarrow aTb$ (using rule $S \rightarrow aTb$)
- $\Rightarrow aTab$ (using rule $T \rightarrow Ta$)
- $\Rightarrow aab$ (using rule $T \rightarrow \varepsilon$)

**PDA Example**

<table>
<thead>
<tr>
<th>State</th>
<th>Input</th>
<th>Stack</th>
<th>Equiv Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>$q_{start}$</td>
<td>aab</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$q_{loop}$</td>
<td>aab</td>
<td>$S$</td>
<td>$S \rightarrow aTb$</td>
</tr>
<tr>
<td>$q_{loop}$</td>
<td>aab</td>
<td>$aTb$</td>
<td>$S \rightarrow aTb$</td>
</tr>
<tr>
<td>$q_{loop}$</td>
<td>ab</td>
<td>$Tb$</td>
<td></td>
</tr>
<tr>
<td>$q_{loop}$</td>
<td>ab</td>
<td>$Tab$</td>
<td>$T \rightarrow Ta$</td>
</tr>
<tr>
<td>$q_{loop}$</td>
<td>ab</td>
<td>ab$$</td>
<td>$T \rightarrow \varepsilon$</td>
</tr>
<tr>
<td>$q_{loop}$</td>
<td>b</td>
<td>b$$</td>
<td></td>
</tr>
<tr>
<td>$q_{loop}$</td>
<td></td>
<td>$$</td>
<td></td>
</tr>
<tr>
<td>$q_{accept}$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A lang is a CFL iff some PDA recognizes it

⇒ If a language is a CFL, then a PDA recognizes it
  • Convert CFG→PDA

⇐ If a PDA recognizes a language, then it’s a CFL
  • To prove this part: show PDA has an equivalent CFG
**PDA→CFG: Prelims**

Before converting PDA to CFG, modify it so:

1. It has a single accept state, $q_{\text{accept}}$.
2. It empties its stack before accepting.
3. Each transition either pushes a symbol onto the stack (a push move) or pops one off the stack (a pop move), but it does not do both at the same time.

**Important:**
This doesn't change the language recognized by the PDA.
PDA $P \rightarrow$ CFG $G$: Variables

$P = (Q, \Sigma, \Gamma, \delta, q_0, \{q_{\text{accept}}\})$  

variables of $G$ are $\{A_{pq} \mid p, q \in Q\}$

- **Want:** if $P$ goes from state $p$ to $q$ reading input $x$, then some $A_{pq}$ generates $x$

- **So:** For every pair of states $p, q$ in $P$, add variable $A_{pq}$ to $G$

- **Then:** connect the variables together by,
  - **Add rules:** $A_{pq} \rightarrow A_{pr}A_{rq}$, for each state $r$
  - These rules allow grammar to simulate every possible transition
  - (We haven’t added input read/generated terminals yet)

- **To add terminals:** pair up stack pushes and pops (essence of a CFL)
PDA $P \rightarrow$ CFG $G$: Generating Strings

$P = (Q, \Sigma, \Gamma, \delta, q_0, \{q_{accept}\})$

variables of $G$ are $\{A_{pq} \mid p, q \in Q\}$

- **The key**: pair up stack pushes and pops (essence of a CFL)

  if $\delta(p, a, \varepsilon)$ contains $(r, u)$ and $\delta(s, b, u)$ contains $(q, \varepsilon)$,

  put the rule $A_{pq} \rightarrow aA_{rs}b$ in $G$
PDA $P \rightarrow$ CFG $G$: Generating Strings

$P = (Q, \Sigma, \Gamma, \delta, q_0, \{q_{\text{accept}}\})$  
variables of $G$ are $\{A_{pq} \mid p, q \in Q\}$

- **The key**: pair up stack pushes and pops (essence of a CFL)

  if $\delta(p, a, \varepsilon)$ contains $(r, u)$ and $\delta(s, b, u)$ contains $(q, \varepsilon)$,

  put the rule $A_{pq} \rightarrow aA_{rs}b$ in $G$
PDA $P \rightarrow$ CFG $G$: Generating Strings

$P = (Q, \Sigma, \Gamma, \delta, q_0, \{q\text{accept}\})$

variables of $G$ are $\{A_{pq}| p, q \in Q\}$

- The key: pair up stack pushes and pops (essence of a CFL)

if $\delta(p, a, \varepsilon)$ contains $(r, u)$ and $\delta(s, b, u)$ contains $(q, \varepsilon)$,

put the rule $A_{pq} \rightarrow aA_{rs}b$ in $G$
A language is a CFL ⇔ A PDA recognizes it

✓ ⇒ If a language is a CFL, then a PDA recognizes it
  • Convert CFG→PDA

✓ ⇐ If a PDA recognizes a language, then it’s a CFL
  • Convert PDA→CFG
Regular vs Context-Free Languages (and others?)
Is This Diagram “Correct”?  
(What are the statements implied by this diagram?)

1. Every regular language is a CFL

2. Not every CFL is a regular language
How to **Prove** This Diagram “Correct”?

1. Every regular language is a CFL

2. Not every CFL is a regular language
   
   Find a CFL that is not regular
   
   \( \{ 0^n 1^n | n \geq 0 \} \)
   
   - It’s a CFL
   - *Proof:* CFG \( S \rightarrow 0S1 | \varepsilon \)
   - It’s not regular
   - *Proof:* by contradiction using the Pumping Lemma

\( \{ 0^n 1^n | n \geq 0 \} \)
How to **Prove** This Diagram “Correct”? 

1. **Every regular language is a CFL**  
   - For any regular language \( A \), show ...  
   - ... it has a CFG or PDA  

2. **Not every CFL is a regular language**  
   - A regular language is represented by a:  
     - DFA  
     - NFA  
     - Regular Expression  

context-free languages (CFLs)  
regular languages
Regular Languages are CFLs: 3 Ways to Prove

• DFA → CFG or PDA

• NFA → CFG or PDA

• Regular expression → CFG or PDA

See HW 6!

Are there other interesting subsets of CFLs?
Deterministic CFLs and DPDAs
Previously: Generating Strings

Generating strings:
1. **Start** with **start variable**, 
2. **Repeatedly apply** CFG rules to get string (and parse tree)

$$A \to 0A1$$
$$A \to B$$
$$B \to \#$$

$$A \Rightarrow 0A1 \Rightarrow 00A11 \Rightarrow 000A111 \Rightarrow 000B111 \Rightarrow 000\#111$$
Generating vs Parsing

Generating strings:
1. Start with start variable,
2. Repeatedly apply CFG rules
to get string (and parse tree)

In practice, opposite is more interesting:
1. Start with string,
2. Then parse into parse tree

\[
A \rightarrow 0A1 \\
A \rightarrow B \\
B \rightarrow \#
\]

\[
A \Rightarrow 0A1 \Rightarrow 00A11 \Rightarrow 000A111 \Rightarrow 000B111 \Rightarrow 000\#111
\]
Generating vs Parsing

• In practice, parsing a string more important than generating one
  • E.g., a compiler (first) parses source code into a parse tree
  • (Actually, any program with string inputs must first parse it)
Previously: Example CFG $\xrightarrow{}$ PDA

Example Derivation using CFG:
- $S \xrightarrow{} aTb$ (using rule $S \xrightarrow{} aTb$)
- $\Rightarrow aTab$ (using rule $T \xrightarrow{} Ta$)
- $\Rightarrow aab$ (using rule $T \xrightarrow{} \varepsilon$)

PDA Example

<table>
<thead>
<tr>
<th>State</th>
<th>Input</th>
<th>Stack</th>
<th>Equiv Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>$q_{\text{start}}$</td>
<td>aab</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$q_{\text{loop}}$</td>
<td>aab</td>
<td>$S$</td>
<td></td>
</tr>
<tr>
<td>$q_{\text{loop}}$</td>
<td>aab</td>
<td>$aTb$</td>
<td>$S \xrightarrow{} aTb$</td>
</tr>
<tr>
<td>$q_{\text{loop}}$</td>
<td>ab</td>
<td>$Tb$</td>
<td></td>
</tr>
<tr>
<td>$q_{\text{loop}}$</td>
<td>ab</td>
<td>$Tab$</td>
<td>$T \xrightarrow{} Ta$</td>
</tr>
<tr>
<td>$q_{\text{loop}}$</td>
<td>ab</td>
<td>$ab$</td>
<td></td>
</tr>
<tr>
<td>$q_{\text{loop}}$</td>
<td>b</td>
<td>$b$</td>
<td></td>
</tr>
<tr>
<td>$q_{\text{loop}}$</td>
<td></td>
<td>$$</td>
<td></td>
</tr>
<tr>
<td>$q_{\text{accept}}$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This Machine is parsing:
1. **Start** with (input) string,
2. **Find rules** that **generate** string
Generating vs Parsing

• In practice, parsing a string more important than generating one
  • E.g., a compiler (first step) parses source code into a parse tree
  • (Actually, any program with string inputs must first parse it)

• But: the PDAs we’ve seen are non-deterministic (like NFAs)
Previously: (Nondeterministic) PDA

\[ S \rightarrow a Tb \mid b \]
\[ T \rightarrow Ta \mid \varepsilon \]

\[ \varepsilon, S \rightarrow b \] \quad \[ \varepsilon, \varepsilon \rightarrow T \] \quad \[ \varepsilon, \varepsilon \rightarrow a \]

\[ \varepsilon, T \rightarrow a \] \quad \[ \varepsilon, \varepsilon \rightarrow T \]

\[ \varepsilon, S \rightarrow b \] \quad \[ \varepsilon, T \rightarrow \varepsilon \] \quad \[ a, a \rightarrow \varepsilon \]
\[ b, b \rightarrow \varepsilon \]

This PDA nondeterministically “tries all grammar rules at once”

A parser implementation can’t do this!
Generating vs Parsing

• In practice, **parsing** a string more important than **generating** one
  • E.g., a **compiler** (first step) parses source code into a parse tree
  • (Actually, *any* program with string inputs must first parse it)

• But: the PDAs we’ve seen are **non-deterministic** (like NFAs)

• Compiler’s parsing algorithm must be **deterministic**

• **So:** to model parsers, we need a **Deterministic PDA** (DPDA)
DPDA: Formal Definition

A deterministic pushdown automaton is a 6-tuple \((Q, \Sigma, \Gamma, \delta, q_0, F)\), where \(Q\), \(\Sigma\), \(\Gamma\), and \(F\) are all finite sets, and

1. \(Q\) is the set of states,
2. \(\Sigma\) is the input alphabet,
3. \(\Gamma\) is the stack alphabet,
4. \(\delta: Q \times \Sigma \times \Gamma \rightarrow (Q \times \Gamma) \cup \{\emptyset\}\) is the transition function,
5. \(q_0 \in Q\) is the start state, and
6. \(F \subseteq Q\) is the set of accept states.

A pushdown automaton is a 6-tuple

1. \(Q\) is the set of states,
2. \(\Sigma\) is the input alphabet,
3. \(\Gamma\) is the stack alphabet,
4. \(\delta: Q \times \Sigma \times \Gamma \rightarrow \mathcal{P}(Q \times \Gamma)\)
5. \(q_0 \in Q\) is the start state, and
6. \(F \subseteq Q\) is the set of accept states.

The language of a DPDA is called a deterministic context-free language.

Difference: DPDA has only one possible action, for any given state, input, and stack op (similar to DFA vs NFA).

Must take into account \(\varepsilon\) reads or stack ops! E.g., if \(\delta(q, a, X)\) does “something”, then \(\delta(q, \varepsilon, X)\) must “do nothing”.

“do nothing”
DPDAs are **Not** Equivalent to PDAs!

- A PDA can non-deterministically "try all rules" (abandoning failed attempts);
- A DPDA must **choose one rule at each step**! (can't go back after reading input!)

\[
\begin{align*}
R & \rightarrow S | T \\
S & \rightarrow aSb | ab \\
T & \rightarrow aTbb | abb
\end{align*}
\]

### Parsing
- Parsing = deriving reversed: **start** with string, **end** with parse tree

### Example Parsing
- **Used S rule**
- **Used T rule**

When parsing this string, when does it know which rule was used, \( S \) or \( T \)?

Choosing “correct” rule depends on rest of the input!

PDAs recognize CFLs, but **DPDAs only recognize DCFLs**! (a **subset** of CFLs)
Subclasses of CFLs

Unambiguous CFLs / PDAs

DCFLs

Programming language parsers / compilers are ideally in here

Unambiguous Grammars

\( LL(k) \) \hspace{1cm} \( LR(k) \)
\( LL(1) \) \hspace{1cm} \( LR(1) \)
\( LALR(1) \)
\( SLR \)
\( LL(0) \) \hspace{1cm} \( LR(0) \)

Ambiguous Grammars

All CFLS
Compiler Stages

A program string (chars) (e.g., \( a : = ( 5 + 3 ) ; \ldots \))

DFAs (recognizing regular languages) in here!

Lexer

Program “words”
(e.g., \texttt{ID(a) ASSIGN LPAREN NUM(5) PLUS NUM(3) RPAREN SEMI \ldots})
A Lexer Implementation

```c
{%
/* C Declarations: */
#include "tokens.h" /* definitions of IF, ID, NUM, ... */
#include "errormsg.h"
union {int ival; string sval; double fval;} yylval;
int charPos=1;
#define ADJ (EM_tokPos=charPos, charPos+=yyleng)
%
/* Lex Definitions: */
digits [0-9]+ %
/* Regular Expressions and Actions: */
if
   [a-z][a-z0-9]*
   {ADJ; return IF;}
   {ADJ; yylval.sval=String(yytext);
    return ID;}
{digits}
   {ADJ; yylval.ival=atoi(yytext);
    return NUM;}
({digits}"."[0-9]*)|([0-9]"."{digits})
   {ADJ;
    yylval.fval=atof(yytext);
    return REAL;}
("-"[a-z]*"\n")|(" "|"\n"|"\t").*
   {ADJ;}
{ADJ; EM_error("illegal character");}
%
```

Remember our analogy:
- DFAs are like programs
- All possible DFA tuples is like a programming language

This DFA is a real program!

A “lex” tool converts the program:
- from “DFA Lang” ...
- to an equivalent one in C!
Compiler Stages

A program (chars) (e.g., \( a := (5 + 3) \); ...)

DFAs (recognizing regular languages) in here!

Lexer

Program “words” (e.g., \( \text{ID}(a) \\text{ASSIGN} \text{LPAREN} \text{NUM}(5) \\text{PLUS} \text{NUM}(3) \text{RPAREN} \text{SEMI} \) ...)

DPDAs (recognizing DCFLs) in here!

Parser

Abstract Syntax tree (AST), i.e., a parse tree!

```
AssignStmt
  a
  OpExp
    NumExp
      5
    Plus
    NumExp
      3
```
A Parser Implementation

```c
{%
int yylex(void);
void yyerror(char *s) { EM_error(EM_tokPos, "%s", s); }
%
%token ID WHILE BEGIN END DO IF THEN ELSE SEMI ASSIGN
%start prog
%
prog: stmlist

stm : ID ASSIGN ID
    | WHILE ID DO stm
    | BEGIN stmlist END
    | IF ID THEN stm
    | IF ID THEN stm ELSE stm

stmlist : stm
    | stmlist SEMI stm

Remember our analogy: CFGs are like programs
This CFG is a real program!
A “yacc” tool converts the program:
- from “CFG Lang” ...
- to an equivalent one in C !
```

Just write the CFG!
DPDAs are **Not** Equivalent to PDAs!

\[
R \rightarrow S | T \\
S \rightarrow aSb \mid ab \\
T \rightarrow aTbb \mid abb
\]

**PDA**: can non-deterministically "try all rules" (abandoning failed attempts); **DPDA**: must choose one rule at each step!

- Parsing = generating reversed:
  - start with string
  - end with parse tree

When parsing reaches this position, does it know which rule, S or T?

Should use S rule

To choose "correct" rule, need to "look ahead" at rest of the input!

PDAs recognize CFLs, but **DPDAs only recognize DCFLs**! (a **subset** of CFLs)
Subclasses of CFLs

2 parser design decisions:
1) Parse from left, or from right
2) choose “look ahead” amount

Programming language parsers / compilers are ideally in here

All CFLs
LL parsing

- L = left-to-right
- L = leftmost derivation

Game: “You’re the Parser”:
Guess which rule applies?
(And how much did you have to “look ahead”?)

1 \( S \rightarrow \text{if } E \text{ then } S \text{ else } S \)
2 \( S \rightarrow \text{begin } S \ L \)
3 \( S \rightarrow \text{print } E \)

if 2 = 3 begin print 1; print 2; end else print 0

4 \( L \rightarrow \text{end} \)
5 \( L \rightarrow ; \ S \ L \)
6 \( E \rightarrow \text{num = num} \)
LL parsing

- $L = \text{left-to-right}$
- $L = \text{leftmost derivation}$

1. $S \rightarrow \text{if } E \text{ then } S \text{ else } S$
2. $S \rightarrow \text{begin } S \ L$
3. $S \rightarrow \text{print } E$
4. $L \rightarrow \text{end}$
5. $L \rightarrow ; \ S \ L$
6. $E \rightarrow \text{num = num}$

if 2 = 3 begin print 1; print 2; end else print 0
LL parsing

- L = left-to-right
- L = leftmost derivation

1. $S \rightarrow \text{if } E \text{ then } S \text{ else } S$
2. $S \rightarrow \text{begin } S \text{ } L$
3. $S \rightarrow \text{print } E$
4. $L \rightarrow \text{end}$
5. $L \rightarrow ; \text{ } S \text{ } L$
6. $E \rightarrow \text{num } = \text{ num}$

if 2 = 3 begin print 1; print 2; end else print 0
LL parsing

- L = left-to-right
- L = leftmost derivation

1. $S \rightarrow \text{if } E \text{ then } S \text{ else } S$
2. $S \rightarrow \text{begin } S \ L$
3. $S \rightarrow \text{print } E$
4. $L \rightarrow \text{end}$
5. $L \rightarrow ; \ S \ L$
6. $E \rightarrow \text{num } = \text{num}$

if 2 = 3 begin print 1; print 2; end else print 0

“Prefix” languages (Scheme/Lisp) are easily parsed with LL parsers (zero lookahead)
LR parsing

- **L** = left-to-right
- **R** = rightmost derivation

1. \( S \rightarrow S ; S \)
2. \( S \rightarrow \text{id} := E \)
3. \( S \rightarrow \text{print}(L) \)
4. \( E \rightarrow \text{id} \)
5. \( E \rightarrow \text{num} \)
6. \( E \rightarrow E + E \)

```plaintext
a := 7;
b := c + (d := 5 + 6, d)
```

When parse is here, can’t determine whether it’s an assign (:=) or addition (+)

Need to **save** input (lookahead) to some memory, like a **stack**! this is a job for a (D)PDA!
LR parsing

- **L** = left-to-right
- **R** = rightmost derivation

\[
\begin{align*}
S & \rightarrow S ; \; S \\
S & \rightarrow id := E \\
S & \rightarrow print( L ) \\
E & \rightarrow id \\
E & \rightarrow num \\
E & \rightarrow E + E
\end{align*}
\]

\[
\begin{align*}
a & := 7 ; \\
\textbf{b} & := c + ( d := 5 + 6 , \; d )
\end{align*}
\]
LR parsing

- \( L = \) left-to-right
- \( R = \) rightmost derivation

\[
\begin{align*}
S & \rightarrow S ; \ S & \quad E & \rightarrow \text{id} \\
S & \rightarrow \text{id} := E & \quad E & \rightarrow \text{num} \\
S & \rightarrow \text{print} \ ( L ) & \quad E & \rightarrow E + E
\end{align*}
\]
LR parsing

- **L** = left-to-right
- **R** = rightmost derivation

```
S → S ; S
S → id := E
E → id
E → num
S → print ( L )
E → E + E
```

---

**Stack**

<table>
<thead>
<tr>
<th>Stack</th>
<th>Input</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>shift</td>
</tr>
<tr>
<td>1 id4</td>
<td></td>
<td>shift</td>
</tr>
<tr>
<td>1 id4 :=6</td>
<td></td>
<td>shift</td>
</tr>
<tr>
<td>1 id4 :=6 num10</td>
<td>7 ; b := c + ( d := 5 + 6 , d )</td>
<td>reduce E → num</td>
</tr>
<tr>
<td>1</td>
<td>a := 7 ; b := c + ( d := 5 + 6 , d )</td>
<td>$</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>shift</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```
LR parsing

- **L** = left-to-right
- **R** = rightmost derivation

1. $S \rightarrow S ; S$
2. $S \rightarrow \text{id := } E$
3. $S \rightarrow \text{print (} L \text{)}$
4. $E \rightarrow \text{id}$
5. $E \rightarrow \text{num}$
6. $E \rightarrow E + E$

**Stack**

<table>
<thead>
<tr>
<th>Stack</th>
<th>Input</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>shift</td>
</tr>
<tr>
<td>id$_4$</td>
<td></td>
<td>shift</td>
</tr>
<tr>
<td>1 id$_4$ :=</td>
<td></td>
<td>shift</td>
</tr>
<tr>
<td>num$_{10}$</td>
<td></td>
<td>reduce $E \rightarrow \text{num}$</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>:= 7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>:= c + ( d := 5 + 6 , d )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>:= c + ( d := 5 + 6 , d )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>:= c + ( d := 5 + 6 , d )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>; b := c + ( d := 5 + 6 , d )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Can determine (rightmost) rule
LR parsing

- L = left-to-right
- R = rightmost derivation

1. $S \rightarrow S ; \ S$
2. $S \rightarrow \text{id} := E$
3. $S \rightarrow \text{print} \ ( \ L )$
4. $E \rightarrow \text{id}$
5. $E \rightarrow \text{num}$
6. $E \rightarrow E + E$

<table>
<thead>
<tr>
<th>Stack</th>
<th>Input</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>a := 7 ; b := c + ( d := 5 + 6 , d ) $</td>
<td>shift</td>
</tr>
<tr>
<td>1 id_4</td>
<td>:= 7 ; b := c + ( d := 5 + 6 , d ) $</td>
<td>shift</td>
</tr>
<tr>
<td>1 id_4 := 6</td>
<td>:= c + ( d := 5 + 6 , d ) $</td>
<td>shift</td>
</tr>
<tr>
<td>1 id_4 := 6 num_10</td>
<td>:= c + ( d := 5 + 6 , d ) $</td>
<td>reduce $E \rightarrow \text{num}$</td>
</tr>
<tr>
<td>1 id_4 := 6 E_11</td>
<td>; b := c + ( d := 5 + 6 , d ) $</td>
<td>reduce $S \rightarrow \text{id} := E$</td>
</tr>
</tbody>
</table>
LR parsing

- L = left-to-right
- R = rightmost derivation

\[ S \rightarrow S \; ; \; S \]
\[ S \rightarrow \text{id} := E \]
\[ E \rightarrow \text{id} \]
\[ E \rightarrow \text{num} \]
\[ S \rightarrow \text{print} \ ( L ) \]
\[ E \rightarrow E + E \]

<table>
<thead>
<tr>
<th>Stack</th>
<th>Input</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>a := 7 ; b := c + ( d := 5 + 6 , d ) $</td>
<td>shift</td>
</tr>
<tr>
<td>1 id4</td>
<td>:= 7 ; b := c + ( d := 5 + 6 , d ) $</td>
<td>shift</td>
</tr>
<tr>
<td>1 id4 := 6</td>
<td>7 ; b := c + ( d := 5 + 6 , d ) $</td>
<td>shift</td>
</tr>
<tr>
<td>1 id4 := 6</td>
<td>; b := c + ( d := 5 + 6 , d ) $</td>
<td>reduce E → num</td>
</tr>
<tr>
<td>1 id4 := 6</td>
<td>; b := c + ( d := 5 + 6 , d ) $</td>
<td>reduce S → id := E</td>
</tr>
<tr>
<td>1 E11</td>
<td>; b := c + ( d := 5 + 6 , d ) $</td>
<td>shift</td>
</tr>
<tr>
<td>1 S2</td>
<td>; b := c + ( d := 5 + 6 , d ) $</td>
<td></td>
</tr>
</tbody>
</table>
To learn more, take a Compilers Class!

This phase needs computation that goes beyond CFLs