CS323: Compilers

Spring 2023

Week 4: Parsers - Top-Down Parsing (table-driven approach contd. and background concepts), Bottom-up parsing (use of goto and action tables)

Parsing using stack-based model

- How do we use the Parse Table constructed?

Top-Down Parsing - Example

`string: xacc$`

Stack	Rem. Input	Action
?	xacc$	?

What do you put on the stack?

Top-Down Parsing - Example

string: xacc$

Stack	Rem. Input	Action
?	xacc$	?

What do you put on the stack? – strings that you derive

Top-Down Parsing - Example

string: xacc$

Stack*	**Rem. Input**	**Action**
S	xacc$	?

Top-down parsing. So, start with S.

* Stack top is on the left-side (first symbol) of the column

Top-Down Parsing - Example

string: xacc$

Stack*	**Rem. Input**	**Action**
S	xacc$	?

Top-down parsing. So, start with S.

What action do you take when stack-top has symbol S and the string to be matched has terminal x in front?

* Stack top is on the left-side (first symbol) of the column

Top-Down Parsing - Example

string: xacc$

Stack*	**Rem. Input**	**Action**
S	xacc$	Predict(1) S->ABc$

Top-down parsing. So, start with S.

What action do you take when stack-top has symbol S and the string to be matched has terminal x in front? – consult parse table

	x	y	a	b	c	$
S	1	1			1	
A	2	3			4	
B				5	6	

* Stack top is on the left-side (first symbol) of the column

Top-Down Parsing - Example

string: xacc$

Stack*	Rem. Input	Action
S	xacc$	Predict(1) S->ABc$
ABc$		

	x	y	a	b	c	$
S	1	1			1	
A	2	3			4	
B				5	6	

* Stack top is on the left-side (first symbol) of the column

Top-Down Parsing - Example

string: xacc$

Stack*	**Rem. Input**	**Action**
S	xacc$	Predict(1) S->ABc$
ABc$	xacc$	

What action do you take when stack-top has symbol A and the string to be matched has terminal x in front? – consult parse table

	x	y	a	b	c	$
S	1	1			1	
A	2	3			4	
B				5	6	

* Stack top is on the left-side (first symbol) of the column

Top-Down Parsing - Example

string: xacc$

Stack*	**Rem. Input**	**Action**
S	xacc$	Predict(1) S->ABc$
ABc$	xacc$	Predict(2) A->xaA

What action do you take when stack-top has symbol A and the string to be matched has terminal x in front? – consult parse table

	x	y	a	b	c	$
S	1	1			1	
A	2	3			4	
B				5	6	

* Stack top is on the left-side (first symbol) of the column

Top-Down Parsing - Example

string: xacc$

Stack*	**Rem. Input**	**Action**
S	xacc$	Predict(1) S->ABc$
ABc$	xacc$	Predict(2) A->xaA
xaABc$		

	x	y	a	b	c	$
S	1	1			1	
A	2	3			4	
B				5	6	

* Stack top is on the left-side (first symbol) of the column

Top-Down Parsing - Example

string: xacc$

Stack*	Rem. Input	Action
S	xacc$	Predict(1) S->ABc$
ABc$	xacc$	Predict(2) A->xaA
xaABc$	xacc$	?

	x	y	a	b	c	$
S	1	1			1	
A	2	3			4	
B				5	6	

* Stack top is on the left-side (first symbol) of the column

Top-Down Parsing - Example

string: xacc$

Stack*	Rem. Input	Action
S	xacc$	Predict(1) S->ABc$
ABc$	xacc$	Predict(2) A->xaA
xaABc$	xacc$	match(x)

	x	y	a	b	c	$
S	1	1			1	
A	2	3			4	
B				5	6	

* Stack top is on the left-side (first symbol) of the column

Top-Down Parsing - Example

string: xacc$

Stack*	**Rem. Input**	**Action**
S	xacc$	Predict(1) S->ABc$
ABc$	xacc$	Predict(2) A->xaA
xaABc$	xacc$	match(x)
aABc$	acc$	match(a)

	x	y	a	b	c	$
S	1	1			1	
A	2	3			4	
B				5	6	

* Stack top is on the left-side (first symbol) of the column

Top-Down Parsing - Example

string: xacc$

Stack*	**Rem. Input**	**Action**
S	xacc$	Predict(1) S->ABc$
ABc$	xacc$	Predict(2) A->xaA
xaABc$	xacc$	match(x)
aABc$	acc$	match(a)
ABc$	cc$	?

	x	y	a	b	c	$
S	1	1			1	
A	2	3			4	
B				5	6	

* Stack top is on the left-side (first symbol) of the column

Top-Down Parsing - Example

	x	y	a	b	c	$
S	1	1			1	
A	2	3			4	
B				5	6	

string: xacc$

Stack*	**Rem. Input**	**Action**
S	xacc$	Predict(1) S->ABc$
ABc$	xacc$	Predict(2) A->xaA
xaABc$	xacc$	match(x)
aABc$	acc$	match(a)
ABc$	cc$	Predict(4) A->c

* Stack top is on the left-side (first symbol) of the column

Top-Down Parsing - Example

	x	y	a	b	c	$
S	1	1			1	
A	2	3			4	
B				5	6	

string: xacc$

Stack*	Rem. Input	Action
S	xacc$	Predict(1) S->ABc$
ABc$	xacc$	Predict(2) A->xaA
xaABc$	xacc$	match(x)
aABc$	acc$	match(a)
ABc$	cc$	Predict(4) A->c
cBc$		

* Stack top is on the left-side (first symbol) of the column

Top-Down Parsing - Example

	x	y	a	b	c	$
S	1	1			1	
A	2	3			4	
B				5	6	

string: xacc$

Stack*	**Rem. Input**	**Action**
S	xacc$	Predict(1) S->ABc$
ABc$	xacc$	Predict(2) A->xaA
xaABc$	xacc$	match(x)
aABc$	acc$	match(a)
ABc$	cc$	Predict(4) A->c
cBc$	cc$	?

* Stack top is on the left-side (first symbol) of the column

Top-Down Parsing - Example

	x	y	a	b	c	$
S	1	1			1	
A	2	3			4	
B				5	6	

string: xacc$

Stack*	Rem. Input	Action
S	xacc$	Predict(1) S->ABc$
ABc$	xacc$	Predict(2) A->xaA
xaABc$	xacc$	match(x)
aABc$	acc$	match(a)
ABc$	cc$	Predict(4) A->c
cBc$	cc$	match(c)

* Stack top is on the left-side (first symbol) of the column

Top-Down Parsing - Example

	x	y	a	b	c	$
S	1	1			1	
A	2	3			4	
B				5	6	

string: xacc$

Stack*	**Rem. Input**	**Action**
S	xacc$	Predict(1) S->ABc$
ABc$	xacc$	Predict(2) A->xaA
xaABc$	xacc$	match(x)
aABc$	acc$	match(a)
ABc$	cc$	Predict(4) A->c
cBc$	cc$	match(c)
Bc$	c$	?

* Stack top is on the left-side (first symbol) of the column

Top-Down Parsing - Example

	x	y	a	b	c	$
S	1	1			1	
A	2	3			4	
B				5	6	

string: xacc$

Stack*	**Rem. Input**	**Action**
S	xacc$	Predict(1) S->ABc$
ABc$	xacc$	Predict(2) A->xaA
xaABc$	xacc$	match(x)
aABc$	acc$	match(a)
ABc$	cc$	Predict(4) A->c
cBc$	cc$	match(c)
Bc$	c$	Predict(6) B->λ

* Stack top is on the left-side (first symbol) of the column

Top-Down Parsing - Example

	x	y	a	b	c	$
S	1	1			1	
A	2	3			4	
B				5	6	

string: xacc$

Stack*	**Rem. Input**	**Action**
S	xacc$	Predict(1) S->ABc$
ABc$	xacc$	Predict(2) A->xaA
xaABc$	xacc$	match(x)
aABc$	acc$	match(a)
ABc$	cc$	Predict(4) A->c
cBc$	cc$	match(c)
Bc$	c$	Predict(6) B->λ
c$		

* Stack top is on the left-side (first symbol) of the column

Top-Down Parsing - Example

	x	y	a	b	c	$
S	1	1			1	
A	2	3			4	
B				5	6	

string: xacc$

Stack*	**Rem. Input**	**Action**
S	xacc$	Predict(1) S->ABc$
ABc$	xacc$	Predict(2) A->xaA
xaABc$	xacc$	match(x)
aABc$	acc$	match(a)
ABc$	cc$	Predict(4) A->c
cBc$	cc$	match(c)
Bc$	c$	Predict(6) B->λ
c$	c$	?

* Stack top is on the left-side (first symbol) of the column

Top-Down Parsing - Example

	x	y	a	b	c	$
S	1	1			1	
A	2	3			4	
B				5	6	

string: xacc$

Stack*	**Rem. Input**	**Action**
S	xacc$	Predict(1) S->ABc$
ABc$	xacc$	Predict(2) A->xaA
xaABc$	xacc$	match(x)
aABc$	acc$	match(a)
ABc$	cc$	Predict(4) A->c
cBc$	cc$	match(c)
Bc$	c$	Predict(6) B->λ
c$	c$	match(c)

* Stack top is on the left-side (first symbol) of the column

Top-Down Parsing - Example

	x	y	a	b	c	$
S	1	1			1	
A	2	3			4	
B				5	6	

string: xacc$

Stack*	**Rem. Input**	**Action**
S	xacc$	Predict(1) S->ABc$
ABc$	xacc$	Predict(2) A->xaA
xaABc$	xacc$	match(x)
aABc$	acc$	match(a)
ABc$	cc$	Predict(4) A->c
cBc$	cc$	match(c)
Bc$	c$	Predict(6) B->λ
c$	c$	match(c)
$	$	Done!

* Stack top is on the left-side (first symbol) of the column

Identifying LL(1) Grammar

- What we saw was an example of LL(1) Grammar
  - Scan input **L**eft-to-right, produce **L**eft-most derivation with 1 symbol look-ahead

Identifying LL(1) Grammar

- What we saw was an example of LL(1) Grammar
  - Scan input **L**eft-to-right, produce **L**eft-most derivation with 1 symbol look-ahead
- Not all Grammars are LL(1)

A Grammar is LL(1) iff for a production A -> α | β, where

α and β are distinct:

1. For no terminal `a` do both α and β derive strings beginning with `a` (i.e. no common prefix)
2. At most one of α and β can derive an empty string
3. If $\beta \overset{*}{\Rightarrow} \epsilon$, then α does not derive any string beginning with a terminal in `Follow(A)`. If $\alpha \overset{*}{\Rightarrow} \epsilon$, then β does not derive any string beginning with a terminal in `Follow(A)`

Example (Left Factoring)

- Consider

  <stmt> → if <expr> then <stmt list> endif

  <stmt> → if <expr> then <stmt list> else <stmt list> endif

- This is not LL(1) (why?)
- We can turn this in to

  <stmt> → if <expr> then <stmt list> <if suffix>

  <if suffix> → endif

  <if suffix> → else <stmt list> endif

Example (Left Factoring)

- Consider

  <stmt> → if <expr> then <stmt list> endif

  <stmt> → if <expr> then <stmt list> else <stmt list> endif

- This is not LL(1) (why?)
- We can turn this in to

  <stmt> → if <expr> then <stmt list> <if suffix>

  <if suffix> → endif

  <if suffix> → else <stmt list> endif

Left Factoring

A -> α β | α μ

$\Downarrow$

A -> α N

N -> β

N -> μ

Left recursion

- *Left recursion* is a problem for LL(1) parsers
  - LHS is also the first symbol of the RHS
- Consider:

  $E \rightarrow E + T$

- What would happen with the stack-based algorithm?

Left recursion

- *Left recursion* is a problem for LL(1) parsers
  - LHS is also the first symbol of the RHS
- Consider:

  $E \rightarrow E + T$

- What would happen with the stack-based algorithm?

  E
  E + T
  E + T + T
  E + T + T + T

Eliminating Left Recursion

$$A \to A\ \alpha \mid \beta$$

$$\Downarrow$$

$A \to NT$

$N \to \beta$

$T \to \alpha T$

$T \to \lambda$

Eliminating Left Recursion

E -> E + T | T

E -> E1 Etail
E1 -> T
Etail -> + T Etail
Etail -> λ

LL(k) parsers

- Can look ahead more than one symbol at a time
  - *k*-symbol lookahead requires extending first and follow sets
  - 2-symbol lookahead can distinguish between more rules:

    $A \rightarrow ax \mid ay$
- More lookahead leads to more powerful parsers
- What are the downsides?

LL(k)? - Example

1) S - > E
2) E -> (E+E)
3) E -> (E-E)
4) E -> x

string: ((x+x))$

Stack*	**Rem. Input**	**Action**
S	((x+x))$	Predict(1) S->E
E		Predict(2) or Predict(3)?

LL(1)

	(	**+**	**-**	**)**	**x**
S	1				1
E	2,3				4

LL(2)

	((	**+(**	**-(**	**)$**	**(x**
S	1				1
E	2,3				4

Are all grammars LL(k)?

- No! Consider the following grammar:

$$\begin{aligned} S &\rightarrow E \\ E &\rightarrow (E + E) \\ E &\rightarrow (E - E) \\ E &\rightarrow x \end{aligned}$$

- When parsing E, how do we know whether to use rule 2 or 3?
  - Potentially unbounded number of characters before the distinguishing '+' or '–' is found
  - No amount of lookahead will help!

In real languages?

- Consider the if-then-else problem
- `if x then y else z`
- Problem: else is optional
- `if a then if b then c else d`
  - Which `if` does the `else` belong to?
- This is analogous to a "bracket language": $[^i\ ]^j\ (i \geq j)$

$$
\begin{aligned}
S &\rightarrow [\ S\ C \\
S &\rightarrow \lambda \\
C &\rightarrow\ ] \\
C &\rightarrow \lambda
\end{aligned}
$$

[[ ] can be parsed: SSλC or SSCλ
(it's ambiguous!)

Solving the if-then-else problem

- The ambiguity exists at the language level. To fix, we need to define the semantics properly
  - "] matches nearest unmatched ["
  - This is the rule C uses for if-then-else
  - What if we try this?

$$
\begin{aligned}
S &\rightarrow [\ S \\
S &\rightarrow S1 \\
S1 &\rightarrow [\ S1\ ] \\
S1 &\rightarrow \lambda
\end{aligned}
$$

This grammar is still not LL(1) (or LL(k) for any k!)

Two possible fixes

- If there is an ambiguity, prioritize one production over another
  - e.g., if C is on the stack, always match "]" before matching "λ"

$$
\begin{aligned}
S &\rightarrow [\ S\ C \\
S &\rightarrow \lambda \\
C &\rightarrow\ ] \\
C &\rightarrow \lambda
\end{aligned}
$$

- Another option: change the language!
  - e.g., all if-statements need to be closed with an endif

$$
\begin{aligned}
S &\rightarrow \text{if } S\ E \\
S &\rightarrow \text{other} \\
E &\rightarrow \text{else } S \text{ endif} \\
E &\rightarrow \text{endif}
\end{aligned}
$$

Parsing if-then-else

- What if we don't want to change the language?
  - C does not require { } to delimit single-statement blocks
- To parse if-then-else, *we need to be able to look ahead at the entire rhs of a production* before deciding which production to use
  - In other words, we need to determine how many "]" to match before we start matching "["s
- *LR parsers* can do this!

Bottom-up Parsing

- More general than top-down parsing
- Used in most parser-generator tools
- Need not have left-factored grammars (i.e. can have left recursion)
- E.g. can work with the bracket language