CS251: Introduction to Language Processing

Top-Down Parsing

Vishwesh Jatala

Department of CSE

Indian Institute of Technology Bhilai

vishwesh@iitbhilai.ac.in

2023-24-M

Acknowledgement

- Today's slides are modified from that of
  - *Stanford University:*
    - *https://web.stanford.edu/class/archive/cs/cs143/cs143.1128/*

LL(1) Tables with ε

Num → Sign Digits

Sign → + | - | ε

Digits → Digit More

More → Digits | ε

Digit → 0 | 1 | … | 9

LL(1) Tables with ε

Num → Sign Digits

Sign → + | - | ε

Digits → Digit More

More → Digits | ε

Digit → 0 | 1 | … | 9

	+	-	#	$
Num				
Sign				
Digits				
More				
Digit				

LL(1) Tables with ε

Num → **Sign Digits**
Sign → **+** | **-** | **ε**
Digits → **Digit More**
More → **Digits** | **ε**
Digit → **0** | **1** | … | **9**

Num		Sign		Digit		Digits		More	
+	-	+	-	0	5	0	5	0	5
0	5	ε		1	6	1	6	1	6
1	6			2	7	2	7	2	7
2	7			3	8	3	8	3	8
3	8			4	9	4	9	4	9
4	9							ε	

	+	-	#	$
Num				
Sign				
Digits				
More				
Digit				

LL(1) Tables with ε

Num	→ Sign Digits
Sign	→ + | - | ε
Digits	→ Digit More
More	→ Digits | ε
Digit	→ 0 | 1 | … | 9

Num		Sign		Digit		Digits		More	
+	-	+	-	0	5	0	5	0	5
0	5	ε		1	6	1	6	1	6
1	6			2	7	2	7	2	7
2	7			3	8	3	8	3	8
3	8			4	9	4	9	4	9
4	9							ε	

	+	-	#	$
Num				
Sign				
Digits				
More				
Digit				

LL(1) Tables with ε

Num → **Sign Digits**

Sign → + | - | ε

Digits → **Digit More**

More → **Digits** | ε

Digit → 0 | 1 | … | 9

Num		Sign		Digit		Digits		More	
+	-	+	-	0	5	0	5	0	5
0	5	ε		1	6	1	6	1	6
1	6			2	7	2	7	2	7
2	7			3	8	3	8	3	8
3	8			4	9	4	9	4	9
4	9							ε	

	+	-	#	$
Num	Sign Digits	Sign Digits	Sign Digits	
Sign				
Digits				
More				
Digit				

LL(1) Tables with ε

Num → Sign Digits
Sign → + | - | ε
Digits → Digit More
More → Digits | ε
Digit → 0 | 1 | … | 9

Num		Sign		Digit		Digits		More	
+	-	+	-	0	5	0	5	0	5
0	5	ε		1	6	1	6	1	6
1	6			2	7	2	7	2	7
2	7			3	8	3	8	3	8
3	8			4	9	4	9	4	9
4	9							ε	

	+	-	#	$
Num	Sign Digits	Sign Digits	Sign Digits	
Sign				
Digits				
More				
Digit				

LL(1) Tables with ε

Num → Sign Digits
Sign → + | - | ε
Digits → Digit More
More → Digits | ε
Digit → 0 | 1 | … | 9

Num		Sign		Digit		Digits		More	
+	-	+	-	0	5	0	5	0	5
0	5	ε		1	6	1	6	1	6
1	6			2	7	2	7	2	7
2	7			3	8	3	8	3	8
3	8			4	9	4	9	4	9
4	9							ε	

	+	-	#	$
Num	Sign Digits	Sign Digits	Sign Digits	
Sign	+	-		
Digits				
More				
Digit				

LL(1) Tables with ε

Num → Sign Digits
Sign → + | - | ε
Digits → Digit More
More → Digits | ε
Digit → 0 | 1 | … | 9

Num		Sign		Digit		Digits		More	
+	-	+	-	0	5	0	5	0	5
0	5	ε		1	6	1	6	1	6
1	6			2	7	2	7	2	7
2	7			3	8	3	8	3	8
3	8			4	9	4	9	4	9
4	9							ε	

	+	-	#	$
Num	Sign Digits	Sign Digits	Sign Digits	
Sign	+	-		
Digits				
More				
Digit				

LL(1) Tables with ε

Num → Sign Digits

Sign → + | - | ε

Digits → Digit More

More → Digits | ε

Digit → 0 | 1 | … | 9

Num		Sign		Digit		Digits		More	
+	-	+	-	0	5	0	5	0	5
0	5	ε		1	6	1	6	1	6
1	6			2	7	2	7	2	7
2	7			3	8	3	8	3	8
3	8			4	9	4	9	4	9
4	9							ε	

	+	-	#	$
Num	Sign Digits	Sign Digits	Sign Digits	
Sign	+	-		
Digits			Digits More	
More				
Digit				

LL(1) Tables with ε

Num → **Sign Digits**
Sign → + | - | ε
Digits → **Digit More**
More → **Digits** | ε
Digit → 0 | 1 | … | 9

Num		Sign		Digit		Digits		More	
+	-	+	-	0	5	0	5	0	5
0	5	ε		1	6	1	6	1	6
1	6			2	7	2	7	2	7
2	7			3	8	3	8	3	8
3	8			4	9	4	9	4	9
4	9							ε	

	+	-	#	$
Num	Sign Digits	Sign Digits	Sign Digits	
Sign	+	-		
Digits			Digits More	
More				
Digit				

LL(1) Tables with ε

Num → **Sign Digits**
Sign → + | - | ε
Digits → **Digit More**
More → **Digits** | ε
Digit → 0 | 1 | … | 9

Num		Sign		Digit		Digits		More	
+	-	+	-	0	5	0	5	0	5
0	5	ε		1	6	1	6	1	6
1	6			2	7	2	7	2	7
2	7			3	8	3	8	3	8
3	8			4	9	4	9	4	9
4	9							ε	

	+	-	#	$
Num	Sign Digits	Sign Digits	Sign Digits	
Sign	+	-		
Digits			Digits More	
More			Digits	
Digit				

LL(1) Tables with ε

Num → **Sign Digits**
Sign → **+** | **-** | **ε**
Digits → **Digit More**
More → **Digits** | **ε**

Digit → **0** | **1** | … | **9**

Num		Sign		Digit		Digits		More	
+	-	+	-	0	5	0	5	0	5
0	5	ε		1	6	1	6	1	6
1	6			2	7	2	7	2	7
2	7			3	8	3	8	3	8
3	8			4	9	4	9	4	9
4	9							ε	

	+	-	#	$
Num	Sign Digits	Sign Digits	Sign Digits	
Sign	+	-		
Digits			Digits More	
More			Digits	
Digit				

LL(1) Tables with ε

Num → **Sign Digits**
Sign → **+** | **-** | **ε**
Digits → **Digit More**
More → **Digits** | **ε**

Digit → **0** | **1** | … | **9**

Num		Sign		Digit		Digits		More	
+	-	+	-	0	5	0	5	0	5
0	5	ε		1	6	1	6	1	6
1	6			2	7	2	7	2	7
2	7			3	8	3	8	3	8
3	8			4	9	4	9	4	9
4	9							ε	

	+	-	#	$
Num	Sign Digits	Sign Digits	Sign Digits	
Sign	+	-		
Digits			Digits More	
More			Digits	
Digit			#	

LL(1) Tables with ε

Num → **Sign Digits**
Sign → **+** | **-** | **ε**
Digits → **Digit More**
More → **Digits** | **ε**
Digit → **0** | **1** | … | **9**

Num	Sign	Digit	Digits	More
+ -	+ -	0 5	0 5	0 5
0 5	ε	1 6	1 6	1 6
1 6		2 7	2 7	2 7
2 7		3 8	3 8	3 8
3 8		4 9	4 9	4 9
4 9				ε

	+	-	#	$
Num	Sign Digits	Sign Digits	Sign Digits	
Sign	+	-		
Digits			Digits More	
More			Digits	
Digit			#	

LL(1) Tables with ε

Num → **Sign Digits**

Sign → + | - | ε

Digits → **Digit More**

More → **Digits** | ε

Digit → 0 | 1 | … | 9

Num		Sign		Digit		Digits		More	
+	-	+	-	0	5	0	5	0	5
0	5	ε		1	6	1	6	1	6
1	6			2	7	2	7	2	7
2	7			3	8	3	8	3	8
3	8			4	9	4	9	4	9
4	9							ε	

	+	-	#	$
Num	Sign Digits	Sign Digits	Sign Digits	
Sign	+	-		
Digits			Digits More	
More			Digits	
Digit			#	

LL(1) Tables with ε

Num → Sign Digits

Sign → + | - | ε

Digits → Digit More

More → Digits | ε

Digit → 0 | 1 | … | 9

Num		Sign		Digit		Digits		More	
+	-	+	-	0	5	0	5	0	5
0	5	ε		1	6	1	6	1	6
1	6			2	7	2	7	2	7
2	7			3	8	3	8	3	8
3	8			4	9	4	9	4	9
4	9							ε	

	+	-	#	$
Num	Sign Digits	Sign Digits	Sign Digits	
Sign	+	-		
Digits			Digits More	
More			Digits	
Digit			#	

LL(1) Tables with ε

Num → Sign Digits

Sign → + | - | ε

Digits → Digit More

More → Digits | ε

Digit → 0 | 1 | … | 9

Num		Sign		Digit		Digits		More	
+	-	+	-	0	5	0	5	0	5
0	5	ε		1	6	1	6	1	6
1	6			2	7	2	7	2	7
2	7			3	8	3	8	3	8
3	8			4	9	4	9	4	9
4	9							ε	

	+	-	#	$
Num	Sign Digits	Sign Digits	Sign Digits	
Sign	+	-	ε	
Digits			Digits More	
More			Digits	
Digit			#	

LL(1) Tables with ε

Num → Sign Digits
Sign → + | - | ε
Digits → Digit More
More → Digits | ε
Digit → 0 | 1 | … | 9

Num		Sign		Digit		Digits		More	
+	-	+	-	0	5	0	5	0	5
0	5	ε		1	6	1	6	1	6
1	6			2	7	2	7	2	7
2	7			3	8	3	8	3	8
3	8			4	9	4	9	4	9
4	9							ε	

	+	-	#	$
Num	Sign Digits	Sign Digits	Sign Digits	
Sign	+	-	ε	
Digits			Digits More	
More			Digits	
Digit			#	

LL(1) Tables with ε

Num → **Sign Digits**

Sign → **+** | **-** | **ε**

Digits → **Digit More**

More → **Digits** | **ε**

Digit → **0** | **1** | … | **9**

Num		Sign		Digit		Digits		More	
+	-	+	-	0	5	0	5	0	5
0	5	ε		1	6	1	6	1	6
1	6			2	7	2	7	2	7
2	7			3	8	3	8	3	8
3	8			4	9	4	9	4	9
4	9							ε	

	+	-	#	$
Num	Sign Digits	Sign Digits	Sign Digits	
Sign	+	-	ε	
Digits			Digits More	
More			Digits	
Digit			#	

LL(1) Tables with ε

Num → **Sign Digits**
Sign → **+** | **-** | **ε**
Digits → **Digit More**
More → **Digits** | **ε**
Digit → **0** | **1** | … | **9**

Num		Sign		Digit		Digits		More	
+	-	+	-	0	5	0	5	0	5
0	5	ε		1	6	1	6	1	6
1	6			2	7	2	7	2	7
2	7			3	8	3	8	3	8
3	8			4	9	4	9	4	9
4	9							ε	

	+	-	#	$
Num	Sign Digits	Sign Digits	Sign Digits	
Sign	+	-	ε	
Digits			Digits More	
More			Digits	ε
Digit			#	

LL(1) Tables with ε

Num → **Sign Digits**
Sign → **+** | **-** | **ε**
Digits → **Digit More**
More → **Digits** | **ε**

Digit → **0** | **1** | … | **9**

Num		Sign		Digit		Digits		More	
+	-	+	-	0	5	0	5	0	5
0	5	ε		1	6	1	6	1	6
1	6			2	7	2	7	2	7
2	7			3	8	3	8	3	8
3	8			4	9	4	9	4	9
4	9							ε	

	+	-	#	$
Num	Sign Digits	Sign Digits	Sign Digits	
Sign	+	-	ε	
Digits			Digits More	
More			Digits	ε
Digit			#	

LL(1) Tables with ε

Num → **Sign Digits**
Sign → **+** | **-** | **ε**
Digits → **Digit More**
More → **Digits** | **ε**
Digit → **0** | **1** | … | **9**

Num	Sign	Digit	Digits	More
+ -	+ -	0 5	0 5	0 5
0 5	ε	1 6	1 6	1 6
1 6		2 7	2 7	2 7
2 7		3 8	3 8	3 8
3 8		4 9	4 9	4 9
4 9				ε

	+	-	#	$
Num	Sign Digits	Sign Digits	Sign Digits	
Sign	+	-	ε	
Digits			Digits More	
More			Digits	ε
Digit			#	

The Final LL(1) Table Algorithm

- Compute FIRST(**A**) and FOLLOW(**A**) for all nonterminals **A**.
- For each rule **A** → ω, for each terminal **t** ∈ FIRST*(ω), set T[**A**, **t**] = ω.
  - Note that **ε** is not a terminal.
- For each rule **A** → ω, if **ε** ∈ FIRST*(ω), set T[**A**, **t**] = ω for each **t** ∈ FOLLOW(**A**).

A Formal Characterization of LL(1)

- A grammar is LL(1) if there are no conflicts in the table.
  - Every entry in the LL(1) table is unique

Exercise

- Construct the LL(1) parser table

  $S \rightarrow SA \mid b \mid \in$

  $A \rightarrow (A) \mid a$

Exercise

- Construct the LL(1) parser table

  $S \rightarrow SA \mid b \mid \in$

  $A \rightarrow (A) \mid a$

	FIRST	FOLLOW
S	{b, ∈, a, (}	{(, a, $}
A	{(, a}	{), (, a, $}

Exercise

- Construct the LL(1) parser table

  $S \rightarrow SA \mid b \mid \in$

  $A \rightarrow (A) \mid a$

	a	b	(	)	$
S	S->SA S->∈	S->b	S->SA S->∈		S->∈
A	A->a		A->(A)		

Exercise

- Construct the LL(1) parser table

  A → **A**`b` | `c`

A Grammar that is Not LL(1)

- Consider the following (left-recursive) grammar:

  A → **A**`b` | `c`
- FIRST(**A**) = {`c`}
-
-

	b	c
A		**A** → **A**`b` **A** → `c`

A Grammar that is Not LL(1)

- Consider the following (left-recursive) grammar:

  A → **A**`b` | `c`
- FIRST(**A**) = {`c`}
-
-

	b	c
A		**A** → **A**`b` **A** → `c`

- **Cannot uniquely predict production!**
- This is called a **FIRST/FIRST conflict.**

Eliminating Left Recursion

- In general, left recursion can be converted into **right recursion** by a mechanical transformation.

- Consider the grammar

$$A \rightarrow A\omega \mid a$$

- This will produce a followed by some number of ω's.

- Can rewrite the grammar as

$$A \rightarrow aB$$

$$B \rightarrow \varepsilon \mid \omega B$$

The Strengths of LL(1)

- LL(1) is straightforward
  - Can be implemented quickly with a table- driven design.
- LL(1) is Fast
  - Can parse in $O(n\ |G|)$ time, where n is the length of the string and $|G|$ is the size of the grammar.

Exercises

- Text book:
  - Example 4.27
  - Example 4.29
  - Example 4.33

Summary

- **Top-down parsing** tries to derive the user's program from the start symbol.
- **Leftmost BFS** is one approach to top-down parsing; it is mostly of theoretical interest.
- **Leftmost DFS** is another approach to top-down parsing that is uncommon in practice.
- **LL(1)** parsing scans from left-to-right, using one token of lookahead to find a leftmost derivation.

  FIRST sets contain terminals that may be the first symbol of a production.

  FOLLOW sets contain terminals that may follow a nonterminal in a production.

  Left recursion cause LL(1) to fail and can be mechanically eliminated in some cases.

Questions?