The CYK Algorithm

Teoria da Computação

Muhsen e Eduardo Seminário 2



The membership problem

– Problem:

- Given a context-free grammar G and a string w
 - $\mathbf{G} = (V, \Sigma, P, S)$ where
 - » V finite set of non-terminal symbols
 - » Σ (the alphabet) finite set of terminal symbols
 - » P finite set of rules
 - » S start symbol (distinguished element of V)
 - » V and Σ are assumed to be disjoint
 - G is used to generate the string of a language
- Question:
 - Is **w** in **L(G**)?



The CYK Algorithm

- J. Cocke
- D. Younger,
- T. Kasami
 - Independently developed an algorithm to answer this question.



The CYK Algorithm Basics

- The Structure of the rules in a Chomsky Normal Form grammar
- Uses a "dynamic programming" or "table-filling algorithm"



Chomsky Normal Form

- Normal Form is described by a set of conditions that each rule in the grammar must satisfy
- Context-free grammar is in CNF if each rule has one of the following forms:
 - $-A \rightarrow BC$ 2 symbols on right side
 - $-A \rightarrow a$, or terminal symbol
 - $-S \rightarrow \epsilon$ empty string

where B, C \in V – {S}



- Each row corresponds to one length of substrings
 - Bottom Row Strings of length 1
 - Second from Bottom Row Strings of length 2

- Top Row - string 'w'

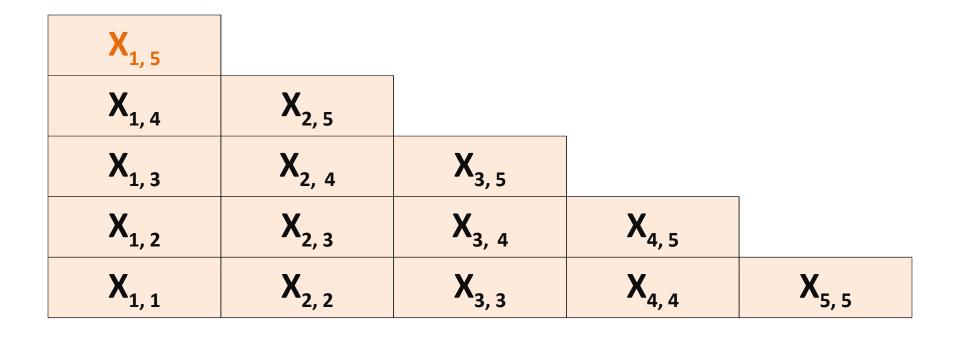


- X_{i, i} is the set of variables A such that
 - $A \rightarrow w_i$ is a production of G

 Compare at most n pairs of previously computed sets:

 $(X_{i,\,i} \text{ , } X_{i+1,\,j} \text{ }), \, (X_{i,\,i+1} \text{ , } X_{i+2,\,j} \text{ }) \ ... \ (X_{i,\,j-1} \text{ , } X_{j,\,j} \text{ })$





w₃

w۵

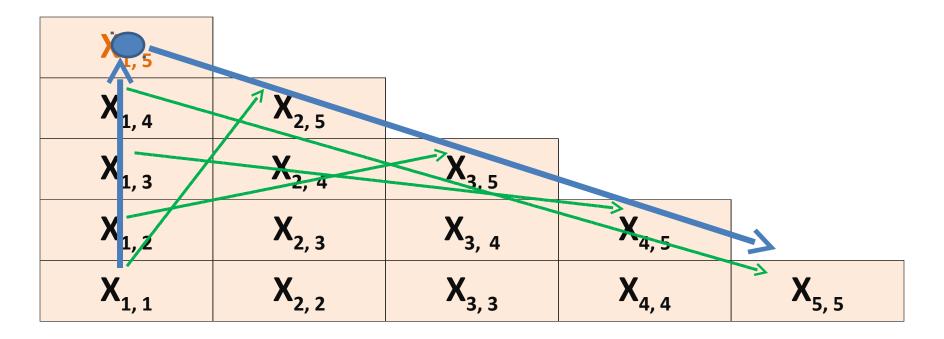
Table for string '**w**' that has length 5

^w2

w₁



w₅



w₁ w₂ w₃ w₄ w₅

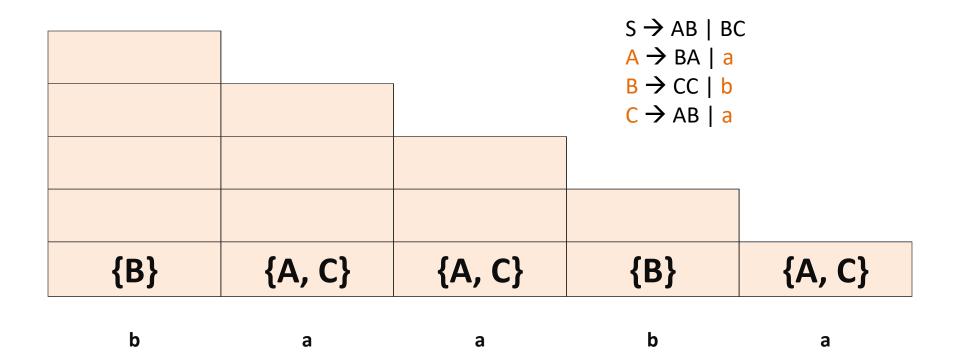
Looking for pairs to compare



Example CYK Algorithm

- Show the CYK Algorithm with the following example:
 - CNF grammar **G**
 - S \rightarrow AB | BC
 - A \rightarrow BA | a
 - $B \rightarrow CC \mid b$
 - C \rightarrow AB | a
 - **w** is baaba
 - Question Is baaba in L(G)?





Calculating the Bottom ROW

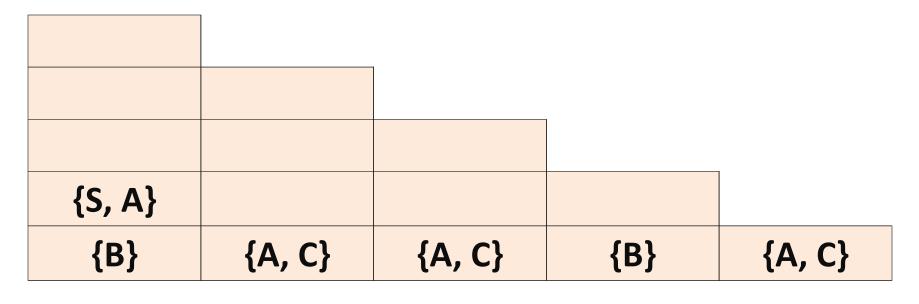


- $X_{1,2} = (X_{i,i}, X_{i+1,j}) = (X_{1,1}, X_{2,2})$
- → {B}{A,C} = {BA, BC}
- Steps:
 - Look for production rules to generate BA or BC
 - There are two: S and A

$$-X_{1,2} = \{S, A\}$$

 $S \rightarrow AB \mid BC$ $A \rightarrow BA \mid a$ $B \rightarrow CC \mid b$ $C \rightarrow AB \mid a$





b	а	а	b	а
---	---	---	---	---

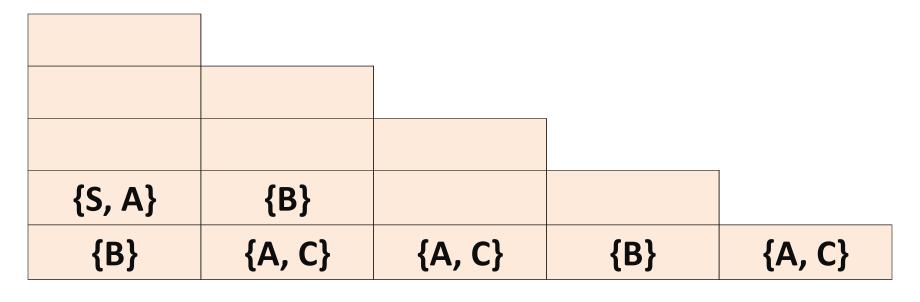


- $X_{2,3} = (X_{i,i}, X_{i+1,j}) = (X_{2,2}, X_{3,3})$
- → {A, C}{A,C} = {AA, AC, CA, CC} = Y
- Steps:
 - Look for production rules to generate Y
 - There is one: B

$$-X_{2,3} = \{B\}$$

 $S \rightarrow AB \mid BC$ $A \rightarrow BA \mid a$ $B \rightarrow CC \mid b$ $C \rightarrow AB \mid a$





b	а	а	b	а
---	---	---	---	---



•
$$X_{3,4} = (X_{i,i}, X_{i+1,j}) = (X_{3,3}, X_{4,4})$$

- \rightarrow {A, C}{B} = {AB, CB} = Y
- Steps:
 - Look for production rules to generate Y
 - There are two: S and C
 - $-X_{3,4} = \{S, C\}$
- $S \rightarrow AB \mid BC$ $A \rightarrow BA \mid a$ $B \rightarrow CC \mid b$ $C \rightarrow AB \mid a$



		_		
{S, A}	{B}	{S, C}		
{B}	{A, C}	{A, C}	{B}	{A, C}

b a	а	b	а
-----	---	---	---



•
$$X_{4,5} = (X_{i,i}, X_{i+1,j}) = (X_{4,4}, X_{5,5})$$

- \rightarrow {B}{A, C} = {BA, BC} = Y
- Steps:
 - Look for production rules to generate Y
 - There are two: S and A

$$-X_{4,5} = \{S, A\}$$

 $S \rightarrow AB \mid BC$ $A \rightarrow BA \mid a$ $B \rightarrow CC \mid b$ $C \rightarrow AB \mid a$



{S, A}	{B}	{S, C}	{S, A}	
{B }	{A, C}	{A, C}	{B}	{A, C}

b	а	а	b	а
---	---	---	---	---



•
$$X_{1,3} = (X_{i,i}, X_{i+1,j}) (X_{i,i+1}, X_{i+2,j})$$

= $(X_{1,1}, X_{2,3}) , (X_{1,2}, X_{3,3})$

- → {B}{B} U {S, A}{A, C}= {BB, SA, SC, AA, AC} = Y
- Steps:
 - Look for production rules to generate Y
 - There are NONE: S and A

$$-X_{1,3} = \emptyset$$

 $S \rightarrow AB \mid BC$ $A \rightarrow BA \mid a$ $B \rightarrow CC \mid b$ $C \rightarrow AB \mid a$

- no elements in this set (empty set)



		_		
Ø				
{S, A}	{B}	{S, C}	{S, A}	
{B}	{A, C}	{A, C}	{B}	{A, C}

b a a b	а
---------	---



•
$$X_{2,4} = (X_{i,i}, X_{i+1,j}) (X_{i,i+1}, X_{i+2,j})$$

= $(X_{2,2}, X_{3,4}), (X_{2,3}, X_{4,4})$

- \rightarrow {A, C}{S, C} U {B}{B}= {AS, AC, CS, CC, BB} = Y
- Steps:
 - Look for production rules to generate Y
 - $S \rightarrow AB \mid BC$ – There is one: B $A \rightarrow BA \mid a$

$$-X_{2,4}=\{B\}$$

 $B \rightarrow CC \mid b$ $C \rightarrow AB \mid a$



		_		
Ø	{B}			
{S, A}	{B}	{S, C}	{S, A}	
{B}	{A, C}	{A, C}	{B}	{A, C}

а

b

а

b



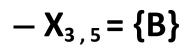
а

•
$$X_{3,5} = (X_{i,i}, X_{i+1,j}) (X_{i,i+1}, X_{i+2,j})$$

= $(X_{3,3}, X_{4,5}), (X_{3,4}, X_{5,5})$

- → {A,C}{S,A} U {S,C}{A,C}
 = {AS, AA, CS, CA, SA, SC, CA, CC} = Y
- Steps:
 - Look for production rules to generate Y
 - There is one: B

 $S \rightarrow AB \mid BC$ $A \rightarrow BA \mid a$ $B \rightarrow CC \mid b$ $C \rightarrow AB \mid a$





		1		
Ø	{B}	{B}		_
{S, A}	{B}	{S, C}	{S, A}	
{B}	{A, C}	{A, C}	{B}	{A, C}

b	а	а	b	а



Final Triangular Table

{S, A, C}	← X _{1, 5}			
Ø	{S, A, C}			
Ø	{B}	{B}		
{S, A}	{B}	{S, C}	{S, A}	
{B}	{A, C}	{A, C}	{B}	{A, C}

а

b

- Table for string 'w' that has length 5

а

b

- The algorithm populates the triangular table



а

Example (Result)

• Is baaba in L(G)?

Yes

We can see the S in the set X_{1n} where 'n' = 5 We can see the table the cell X₁₅ = (S, A, C) then **if S E X₁₅ then baaba E L(G)**





 The CYK Algorithm correctly computes X _{ij} for all i and j; thus w is in L(G) if and only if S is in X_{1n}.



References

- J. E. Hopcroft, R. Motwani, J. D. Ullman, Introduction to Automata Theory, Languages and Computation, Second Edition, Addison Wesley, 2001
- T.A. Sudkamp, Languages and Machines : An Introduction to the Theory of Computer Science, Third Edition, Addison Wesley, 2006



Useful Site

• http://lxmls.it.pt/2015/cky.html

