The CYK Algorithm

- J. Cocke
- D. Younger,
- T. Kasami

The CYK Algorithm

- *Il problema dell'appartenenza*:
  - Problema:
    - Data una grammatica grammar **G** e una stringa **w**
      - **G** = (V, ∑ ,P , S) dove
        - » V insieme finito di variabili
        - » ∑ (alfabeto) insieme finito di simboli terminali-
        - » P insieme finito di produzioni
        - » S simbolo iniziale (elemento distintivo di V)
        - » V e ∑ sono insiemi disgiunti
      - **G** genera un linguaggio, **L**(**G**),
  - Domanda :
    - **w** appartiene al **L**(**G**)?

The CYK Algorithm

– La grammatica è scritta in Chomsky Normal Form
– Viene usata una tecnica chiamata “dynamic programming” o “table-filling algorithm”

Chomsky Normal Form

- *Normal Form è descritta da un insieme di condizioni che ogni regola della grammatica deve soddisfare.*
- Context-free grammar è in CNF, ogni regola ha la seguente forma:
  - $A \rightarrow BC$ al massimo due simboli sul lato destro
  - $A \rightarrow a$ a simbolo terminale
  - $S \rightarrow \lambda$ stringa vuota

  Dove $B, C \in V - \{S\}$

Costruire una Triangular Table

$X_{5,1}$				
$X_{4,1}$	$X_{4,2}$			
$X_{3,1}$	$X_{3,2}$	$X_{3,3}$		
$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$	
$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$	$X_{1,5}$
w_1	w_2	w_3	w_4	w_5

Tavola per una stringa ‘**w**’ che ha lunghezza 5

Esempio CYK Algorithm

- Prendiamo la seguente grammatica:
  - CNF grammatica **G**
    - S → AB | BC
    - A → BA | a
    - B → CC | b
    - C → AB | a
  - **w** sia baaba
  - E' **baaba** in L(G)?

Constructing The Triangular Table

S → AB | BC
A → BA | a
B → CC | b
C → AB | a

{B}	**{A, C}**	**{A, C}**	**{B}**	**{A, C}**
b	**a**	**a**	**b**	**a**

Calcolare la riga più bassa

Costruire la Triangular Table

- $\mathbf{X_{2,1} = (X_{1,1}\ ,X_{1,2})}$
- ➔ {B}{A,C} = {BA, BC}
- Step:
  - Trovare, se esistono, le regole che producono BA or BC
  - Sono due : S e A
  - $\mathbf{X_{2,1} = \{S, A\}}$

S → AB | BC
A → BA | a
B → CC | b
C → AB | a

{S, A}				
{B}	**{A, C}**	**{A, C}**	**{B}**	**{A, C}**
b	**a**	**a**	**b**	**a**

Constructing The Triangular Table

- $\mathbf{X_{2,2} = (X_{1,2}\ , X_{1,3})}$
- ➔ {A, C}{A,C} = {AA, AC, CA, CC} = Y
- Step:
  - Trovare, se esistono, le regole che producono Y
  - Esiste una : B
  - $\mathbf{X_{2,2} = \{B\}}$

S → AB | BC
A → BA | a
B → CC | b
C → AB | a

{S, A}	{B}			
{B}	{A, C}	{A, C}	{B}	{A, C}
b	a	a	b	a

- $\mathbf{X_{2,3} = (X_{1,3}\ , X_{1,4})}$
- ➔ {A, C}{B} = {AB, CB} = Y
- Steps:
  - Trovare, se esistono, le regole che producono Y
  - sono: S e C
  - $\mathbf{X_{2,3} = \{S, C\}}$

S → AB | BC
A → BA | a
B → CC | b
C → AB | a

{S, A}	{B}	{S, C}		
{B}	{A, C}	{A, C}	{B}	{A, C}
b	**a**	**a**	**b**	**a**

- $\mathbf{X_{2,4} = (X_{1,4}\ , X_{1,5})}$
- ➔ {B}{A, C} = {BA, BC} = Y
- Steps:
  - Trovare, se esistono, le regole che producono Y
  - Cono: S and A
  - $\mathbf{X_{2,4} = \{S, A\}}$

S → AB | BC
A → BA | a
B → CC | b
C → AB | a

Constructing The Triangular Table

{S, A}	**{B}**	**{S, C}**	**{S, A}**	
{B}	**{A, C}**	**{A, C}**	**{B}**	**{A, C}**
b	**a**	**a**	**b**	**a**

- $\mathbf{X_{3,1} = (X_{1,1}, X_{2,2}), (X_{2,1}, X_{1,3})}$
- ➔ {B}{B} **U** {S, A}{A, C}= {BB, SA, SC, AA, AC} = Y
- Steps:
  - Trovare, se esistono, le regole che producono Y
  - Nessuna
  - $\mathbf{X_{3,1} = \emptyset}$
  - **Nessun elemento in questo insieme**

S → AB | BC
A → BA | a
B → CC | b
C → AB | a

Ø				
{S, A}	**{B}**	**{S, C}**	**{S, A}**	
{B}	**{A, C}**	**{A, C}**	**{B}**	**{A, C}**
b	**a**	**a**	**b**	**a**

- $\mathbf{X_{3,2} = (X_{1,2}\ , X_{2,3})\ , (X_{2,2}\ , X_{1,3})}$
- ➔ {A, C}{S, C} **U** {B}{B}= {AS, AC, CS, CC, BB} = Y
- Step:
  - Trovare, se esistono, le regole che producono Y
  - una: B
  - $\mathbf{X_{2,4} = \{B\}}$

S → AB | BC
A → BA | a
B → CC | b
C → AB | a

Ø	**{B}**			
{S, A}	**{B}**	**{S, C}**	**{S, A}**	
{B}	**{A, C}**	**{A, C}**	**{B}**	**{A, C}**
b	**a**	**a**	**b**	**a**

- $\mathbf{X_{3,3} = (X_{1,3}, X_{2,4}), (X_{2,3}, X_{1,5})}$
- ➔ {A,C}{S,A} **U** {S,C}{A,C}

  = {AS, AA, CS, CA, SA, SC, CA, CC} = Y
- Step:
  - Trovare, se esistono, le regole che producono Y
  - una: B
  - $\mathbf{X_{3,5} = \{B\}}$

S → AB | BC
A → BA | a
B → CC | b
C → AB | a

Ø	{B}	{B}		
{S, A}	{B}	{S, C}	{S, A}	
{B}	{A, C}	{A, C}	{B}	{A, C}
b	a	a	b	a

- $\mathbf{X_{4,1} = (X_{1,1}, X_{3,2}), (X_{2,1}, X_{2,3})}$
- $\mathbf{(X_{3,1}, X_{1,4})}$
- Step:
  - Trovare, se esistono, le regole che producono Y
  - una: B
  - $\mathbf{X_{4,1} = \{?\}}$

S → AB | BC
A → BA | a
B → CC | b
C → AB | a

- $\mathbf{X_{4,2} = (X_{1,2}, X_{3,3}), (X_{2,2}, X_{2,4})}$
- $\mathbf{(X_{3,2}, X_{1,5})}$
- Step:
  - Trovare, se esistono, le regole che producono Y
  - una: B
  - $\mathbf{X_{4,1} = \{?\}}$

S → AB | BC
A → BA | a
B → CC | b
C → AB | a

Finale Triangular Table

{S, A, C}	**← $X_{5, 1}$**			
Ø	**{S, A, C}**			
Ø	**{B}**	**{B}**		
{S, A}	**{B}**	**{S, C}**	**{S, A}**	
{B}	**{A, C}**	**{A, C}**	**{B}**	**{A, C}**
b	**a**	**a**	**b**	**a**

- Tavola per la stringa ‘**w**’ ha lunghezza 5
- The algorithm popola la triangular table

domanda

– sia **G la grammatical CNF**
  - S → AB | BC
  - A → BA | a
  - B → CC | b
  - C → AB | a
– **w** is ababa
– Domanda: **ababa** è in L(G)?

- E’ baaba in L(G)?

Si

Possiamo vedere che S è nell’insieme X_{1n} dove ‘n’= 5

la cella X_{51} = (S, A, C) allora

S Є X_{15} allora baaba Є L(G)

Construire una Triangular Table

- Ogni riga corrisponde a una lunghezza delle sottostringhe.
  - La riga più in basso – Stringhe di lng 1
  - Seconda riga – Stringhe di lng 2

    .

    .
  - Riga più in alto – la stringa ‘w’

Costruire una Triangular Table

- $\mathbf{X_{i,\,i}}$ è l’insieme delle variabili tale che A è A $\rightarrow$ w_i una produzione di G
- Comparare al massimo n coppie di insieme calcolati in precedenza

$$(\mathbf{X_{i,\,i}}\ \mathbf{,}\ \mathbf{X_{i+1,\,j}}),\ (\mathbf{X_{i,\,i+1}}\ \mathbf{,}\ \mathbf{X_{i+2,\,j}})\ \ldots\ (\mathbf{X_{i,\,j-1}}\ \mathbf{,}\ \mathbf{X_{j,\,j}}\ \mathbf{)}$$

teorema

- The CYK Algorithm calcola correttamente $X_{i\,j}$ per tutti i e j; allora w è in L(G) iff S è in X_{1n}.
- Perchè? Spiegazione, esercizio scrivere la dimostrazione.
- Complessità $O(n^3)$.

- **let** the input be a string L consisting of n characters: a_1 ... a_n.
- **let** the grammar contain r nonterminal symbols R_1 ... R_r, with start symbol R_1.
- **let** $P[n,n,r]$ be an array of booleans. Initialize all elements of P to false.
- **for each** s = 1 to n
- **for each** unit production R_v -> a_s
- set $P[1,s,v]$ = true ; Generata la prima riga-
- **for each** L = 2 to n
- **for each** s = 1 to n-L+1
- **for each** p = 1 to L-1
- **for each** production R_a -> R_b R_c
- **if** $P[p,s,b]$ and $P[L-p,s+p,c]$ **then** set $P[L,s,a]$ = true

- **if** $P[n,1,1]$ is true **then** L is member of language
- **else** L is not member of language

{S, A, C}	← $X_{5, 1}$			
Ø	{S, A, C}			
Ø	{B}	{B}		
{S, A}	{B}	{S, C}	{S, A}	
{B}	{A, C}	{A, C}	{B}	{A, C}
b	**a**	**a**	**b**	**a**

$X_{5,1}$	(5)n-L+1=s			
$X_{4,1}$	$X_{4,2}$	(4)n-L+1=s		
$X_{3,1}$	$X_{3,2}$	$X_{3,3}$	(3)n-L+1 =s	
$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$	(2)n-L+1=s
$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$	$X_{1,5}$
w_1	w_2	w_3	w_4	w_5

for each L = 2 to n
    for each s = 1 to n-L+1
        for each p = 1 to L-1
            for each production R_a -> R_b R_c
                if $P[p,s,b]$ and $P[L\text{-}p,s\text{+}p,c]$ **then** set $P[L,s,a]$ = true

$X_{5,1}$				
$X_{4,1}$	$X_{4,2}$			
$X_{3,1}$	$X_{3,2}$	$X_{3,3}$	**L=3, s=1**	
$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$	
$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$	$X_{1,5}$
w_1	w_2	w_3	w_4	w_5

Tavola per una stringa ‘**w**’ che ha lunghezza 5

```
for each L = 2 to n
    for each s = 1 to n-L+1
        for each p = 1 to L-1
            for each production R_a -> R_b R_c
                if P[p,s,b] and P[L-p,s+p,c] then set P[L,s,a] = true
```

L=4, s=2

$X_{5,1}$				
$X_{4,1}$	$X_{4,2}$			
$X_{3,1}$	$X_{3,2}$	$X_{3,3}$		
$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$	
$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$	$X_{1,5}$
w_1	w_2	w_3	w_4	w_5

Tavola per una stringa ‘**w**’ che ha lunghezza 5

for each L = 2 to n

for each s = 1 to n-L+1

for each p = 1 to L-1

for each production $R_a \rightarrow R_b\ R_c$

if $P[p,s,b]$ and $P[L-p,s+p,c]$ then set $P[L,s,a]$ = true

$X_{5,1}$				
$X_{4,1}$	$X_{4,2}$			
$X_{3,1}$	$X_{3,2}$	$X_{3,3}$		
$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$	
$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$	$X_{1,5}$
w_1	w_2	w_3	w_4	w_5

Tavola per una stringa '**w**' che ha lunghezza 5