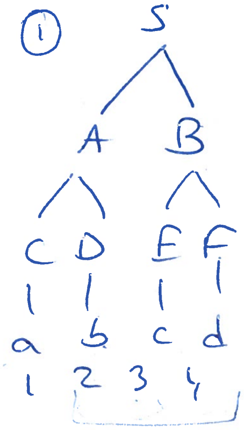
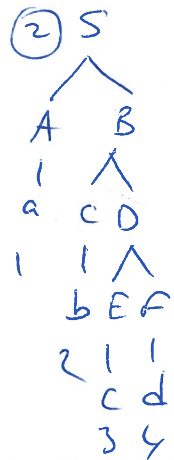
CS 3311
Oct 29, 2012
Monday 1

Last time:

1. $S \to AB$
   $A \to CD$
   $B \to EF$
   $C \to a$
   $D \to b$
   $E \to c$
   $F \to d$

2. $S \to AB$
   $A \to a$
   $B \to CD$
   $C \to b$
   $D \to EF$
   $E \to c$
   $F \to d$

1 

2 

	1	2	3	4
1	C	A	$\emptyset$	S
2		D	$\emptyset$	$\emptyset$
3			E	B
4				F

if $S \in X_{1,4}$ then $w \in L(G)$ else $w \notin L(G)$

	1	2	3	4
1	A	$\emptyset$	$\emptyset$	S
2		C	$\emptyset$	B
3			E	D
4				F

Main question: Given a string w, and a grammar G, is $w \in L(G)$? $S \overset{*}{\Rightarrow} w$?

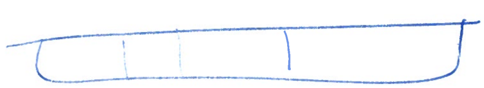
which variable(s) could have initiated the derivation of this part of the string?

w

$S \Rightarrow \underline{\quad A \quad} \Rightarrow$

$\Rightarrow$

$\Rightarrow \underline{\quad A \quad}$



w

Example 4.5.2

$X \to aXb \mid ab$ $\qquad \{a^i b^i \mid i \geq 1\}$

$S \to X$
$X \to aXb \mid ab$ } add new start symbol

---

remove the $S \to X$ chain rule

$S \to aXb \mid ab$ $\qquad S \to A\underbrace{XB}_{T}$
$X \to aXb \mid ab$

Convert to Chomsky Normal form

$S \to AT \mid AB$
$T \to XB$
$X \to AT \mid AB$

$A \to a$
$B \to b$

$w = \underset{1\,2\,3\,4\,5\,6}{aaabbb}$ $\qquad w \in L(G)$? $\qquad n^2$

	1	2	3	4	5	6
1	{A}	∅	∅	∅	∅	S, X
2		A	∅	∅	S, X	T
3			A	S, X	T	∅
4				B	∅	∅
5					B	∅
6						B

diagonal 2
diagonal 1

$S \to AT \mid AB$
$T \to XB$
$X \to AT \mid AB$
$A \to a$
$B \to b$

$X \to aXb \mid ab$

diagonal 2

$X_{1,2}$ $X_{11} \cdot X_{22}$
A A
ϕ

$X_{2,3}$ X_{22} X_{33}
A A
ϕ

$X_{3,4}$ X_{33} X_{44}
A B

$X_{4,5}$ X_{44} X_{55}
B B
ϕ

$X_{5,6}$ X_{55} X_{66}
B B
ϕ

diagonal 3

$X_{1,3}$

X_{11} X_{23}
A $\emptyset$
$\emptyset$

X_{12} X_{33}
$\emptyset$ A
$\emptyset$

$X_{2,4}$

X_{22} X_{34}
A S, X
AS
AX
$\emptyset$

X_{23} X_{44}
$\emptyset$
$\emptyset$

} two possible ways to split into 2

$X_{3,5}$
B

X_{33} $X_{4,5}$
A $\emptyset$
$\emptyset$

$X_{3,4}$ $X_{5,5}$
S, X B
SB
XB } T
~~$\emptyset$~~

} look for these righthand sides

$X_{4,6}$

X_{44} X_{56}
B $\emptyset$
$\emptyset$

X_{45} X_{66}
$\emptyset$
$\emptyset$

diagonal 4

$x_{1,4}$ x_{11} x_{24} x_{12} x_{34} x_{13} x_{44}

A $\emptyset$ $\emptyset$ $\emptyset$

$\emptyset$ $\emptyset$ $\emptyset$

$x_{2,5}$ x_{22} x_{35} x_{23} x_{45} x_{24} x_{55}

A T $\emptyset$ $\emptyset$

S, X $\emptyset$ $\emptyset$

$x_{3,6}$ x_{33} x_{46} x_{34} x_{56} x_{35} x_{66}

A $\emptyset$ S, X $\emptyset$ T B

$\emptyset$ $\emptyset$ $\emptyset$

diagonal 5

$X_{1,5}$

X_{11} X_{25}
A S, X
ϕ

X_{12} X_{35}
ϕ
ϕ

X_{13} X_{45}
ϕ
ϕ

X_{14} X_{55}
ϕ
ϕ

$X_{2,6}$

X_{22} X_{36}
A ϕ
ϕ

X_{23} X_{46}
ϕ
ϕ

X_{24} X_{56}
ϕ
ϕ

X_{25} X_{66}
S, X B
SB
XB T

X_{16}	X_{11} X_{26}	X_{12} X_{36}	X_{13} X_{46}	X_{14} X_{56}	X_{15} X_{66}
	A T	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$
	S, X	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$

$S \in X_{1,6}$ therefore $w = aaabbb \in L(G)$.

8

The CYK Algorithm CS3311 handout

Algorithm 4.6.1 — The CYK algorithm

Given a string of length n, dynamically fills out an $n \times n$ table.

input: context-free grammar $G = (V, \Sigma, P, S)$
string $u = x_1 x_2 \ldots x_n \in \Sigma^*$
private:
X : a table containing sets of variables
step : the index of the "diagonal", the main diagonal is 1, the one above it is 2, and so on.
i: row index (the column index is calculated from it)
k: split position in the string

// Initialize the entire table.
1. initialize all $X_{i,j}$ to $\emptyset$

// Initialize the main diagonal from the rules that derive the terminals of the string.
// The main diagonal (diagonal 1) represents the length 1 substrings.

$O(n)$

2. **for** $i = 1$ to n
  for each variable A
    if there is a rule $A \rightarrow x_{i,i}$ **then**
      $X_{i,i} := X_{i,i} \cup \{A\}$

// Do for each "diagonal."
// *Step* contains the diagonal number. Diagonal n represents the length n substrings.

$O(n) \times O(n)$

$O(n^2)$

card(P)

3. **for** $step = 2$ to n diagonals 2 to n.
  // The cells start from $i, i + step - 1$. 1,2 $6 - 2 + 1 = 5$
  3.1 **for** $i = 1$ to $n - step + 1$
    // i is the row index. It starts at 1. $n - step + 1$ is the last row in this diagonal.
    // For example, the diagonal 2 cells are: 1,2; 2,3; and 3,4.
    // The diagonal 3 cells are: 1,3; 2,4.
    // Do for each split position, k shows the split position.
    3.1.1 **for** $k = i$ to $i + step - 2$
      if there are variables $B \in X_{i,k}$, $C \in X_{k+1,i+step-1}$, and a rule $A \rightarrow BC$ **then**
        $X_{i,i+step-1} = X_{i,i+step-1} \cup \{A\}$

look at right hand sides and fill out.

// If S is in the upper right corner, then the string is in the language. Otherwise, it is not.
4. **if** $S \in X_{1,n}$ **then**
    return TRUE
  else
    return FALSE

CYK runs in $O(n^2)$ time.
number of symbols in a given string.

2^n n^2 is much better than 2^n