Cocke-Younger-Kasami (CYK) algoritmus

Input: Egy környezetfüggetlen $G = \langle T, N, P, S\rangle$ grammatika Chomsky-normálformában adva és egy $u \in T^*$ szó.

Output: $u \overset{?}{\in} L(G)$

Legyen $u = t_1 \dots t_n,\ t_i \in T, n \geq 1$. Legyen A_i a $P_i \in P$ szabály bal-, β_i pedig a jobboldala. ($A_i \in N$, $\beta_i \in T \cup N^2$.)

A CYK algoritmus rekurzíven definiál $H_{i,j},\ 1 \leq i \leq j \leq n$ halmazokat $(j-i)$ szerint növekvő sorrendben.

$$H_{i,i} := \{A_j \mid \beta_j = t_i\}$$

$$H_{i,j} := \{A_k \mid \beta_k \in \bigcup_{h=i}^{j-1} H_{i,h}H_{h+1,j}\} \quad (i < j)$$

$u \in L(G) \iff S \in H_{1,n}$.

Az algoritmus helyességéhez elég belátni, hogy minden $1 \leq i \leq j \leq n$ esetén $H_{i,j} = \{X \in N \mid X \Rightarrow_G^* t_i \cdots t_j\}$.

Ezt $(j-i)$-re vonatkozó teljes indukcióval bizonyítjuk.

$j - i = 0$ esetén világos, hogy t_i éppen $H_{i,i}$ nemterminálisaiból vezethető le.

Tegyük fel, hogy az állítás igaz minden olyan $1 \leq k \leq \ell \leq n$-re, melyre $\ell - k < j - i$ és legyen $1 \leq i < j \leq n$.

Tekintsük egy $X \Rightarrow^* t_i \cdots t_j$ levezetést. Mivel a levezetendő szó legalább 2 hosszú ezért a levezetés első lépése $X \Rightarrow YZ$ valamely $Y, Z \in N$-re.Ekkor létezik egy olyan $i \leq h < j$, melyre $Y \Rightarrow^* t_i \cdots t_h$ és $Z \Rightarrow^* t_{h+1} \cdots t_j$. Mivel $h - i < j - i$ és $j - (h+1) < j - i$ ezért az indukciós feltevés szerint $Y \in H_{i,h}, Z \in H_{h+1,j}$ és így $X \in H_{i,j}$.

Fordítva, ha $X \in H_{i,j}$ $(j > i)$, akkor van olyan $i \leq h < j$ és $Y \in H_{i,h}, Z \in H_{h+1,j}$, melyre $X \to YZ \in P$, azaz $Y \Rightarrow^* t_i \cdots t_h$ és $Z \Rightarrow^* t_{h+1} \cdots t_j$. Ekkor $X \Rightarrow YZ \Rightarrow^* t_i \cdots t_h Z \Rightarrow^* t_i \cdots t_j$.

Példa: A $G = \langle \{S, A, B, C, U, V, W, X, Y, Z\}, \{a, b, c\}, P, S\rangle$ Chomsky normálformájú grammatika esetén CYK algoritmussal döntsük el, hogy az $aabbcc$ szót generálja-e G, ahol P:

$$\begin{aligned}
&S \to AB \,|\, BC\\
&A \to XA \,|\, a\\
&X \to a\\
&C \to YC \,|\, c\\
&Y \to c\\
&B \to UV \,|\, VW\\
&U \to XX\\
&W \to YY\\
&V \to ZZ\\
&Z \to b
\end{aligned}$$

$$\begin{array}{cccccccc}
 & & & & H_{1,n} & & & \\
 & & & H_{1,n-1} & H_{2,n} & & & \\
 & & & & \vdots & & & \\
 & & H_{1,3} & H_{2,4} & & & & \\
 & H_{1,2} & H_{2,3} & & & \cdots & H_{n-1,n} & \\
H_{1,1} & H_{2,2} & & & & & \cdots & H_{n,n} \\
\hline
t_1 & t_2 & & & \cdots & & & t_n
\end{array}$$

$$\begin{array}{cccccccccccc}
 & & & & & \{S\} & & & & & & \\
 & & & & \{S\} & & \{S\} & & & & & \\
 & & & \{B\} & & \{\} & & \{B\} & & & & \\
 & & \{\} & & \{\} & & \{\} & & \{\} & & & \\
 & \{A,U\} & & \{\} & & \{V\} & & \{\} & & \{C,W\} & & \\
\{A,X\} & & \{A,X\} & & \{Z\} & & \{Z\} & & \{Y,C\} & & \{Y,C\} & \\
\hline
a & & a & & b & & b & & c & & c &
\end{array}$$

Mivel $S \in H_{1,6}$, ezért $aabbcc \in L(G)$.