A számításelmélet alapjai I.

8. előadás

előadó: Tichler Krisztián
ktichler@inf.elte.hu

KÖRNYEZETFÜGGETLEN GRAMMATIKÁK

Mostantól környezetfüggetlen grammatikákat fogunk tekinteni.

Emlékeztetőül egy $G = \langle N, T, P, S \rangle$ környezetfüggetlen grammatika szabályai $A \to u$ alakúak, ahol $A \in N, u \in (T \cup N)^*$.

Aktív és elérhető nemterminálisok

Definíció

Egy környezetfüggetlen grammatika egy nemterminálisát **inaktívnak** vagy **nem aktívnak** nevezzük, ha nem vezethető le belőle terminális szó; egyébként **aktívnak** mondjuk.

Definíció

Egy környezetfüggetlen grammatika egy nemterminálisát **nem elérhetőnek** nevezzük, ha nem fordul elő egyetlen mondatformában (azaz, a kezdőszimbólumból levezethető szóban) sem; egyébként **elérhetőnek** mondjuk.

Definíció

Egy környezetfüggetlen grammatika **redukált**, ha minden nemterminálisa aktív és elérhető.

Aktív nemterminálisok meghatározása

Legyen $G = \langle N, T, P, S \rangle$ egy a környezetfüggetlen grammatika

$$A_1 := \{X \mid X \to u \in P, u \in T^*\},$$

$$A_{i+1} := A_i \cup \{X \mid X \to w \in P, w \in (T \cup A_i)^*\}, \quad i = 1, 2, \ldots$$

Nyilván $A_1 \subseteq A_2 \subseteq \cdots \subseteq N$.

Ebből következik, hogy $\exists k \in \mathbb{N}$, amelyre $A_k = A_{k+1}$. Ekkor $\forall \ell \geq k$ esetén $A_\ell = A_k$.

$A := A_k$.

Tétel

A fenti módon előállított A halmaz éppen a G grammatika aktív nemterminálisainak halmaza.

Aktív nemterminálisok meghatározása

Bizonyítás: Legyen $X \in A$, ekkor $X \in A_k$.

Állítás: Minden $0 \leq i \leq k$ esetén létezik $w \in (T \cup A_{k-i})^*$, melyre $X \Rightarrow_G^* w$. ($A_0 := \emptyset$).

i-re vonatkozó teljes indukcióval bizonyítjuk.

$i = 0$-ra nyilvánvaló, hiszen $X \Rightarrow^* X$ és $X \in A_k \subseteq (T \cup A_{k-0})^*$.

Tegyük fel, hogy létezik olyan $w \in (T \cup A_{k-i})^*$, amelyre $X \Rightarrow_G^* w$ teljesül. Ekkor A_{k-i} definíciója szerint w minden nemterminálisából egy lépésben levezethető egy $(T \cup A_{k-(i+1)})^*$-beli szó. w minden nemtermilásából vezessünk el egy ilyen szót, így létezik olyan $w' \in (T \cup A_{k-(i+1)})^*$, melyre $X \Rightarrow_G^* w \Rightarrow_G^* w'$ fennáll.

$i = k$-ra alkalmazva az állítást adódik, hogy A minden nemterminálisa aktív.

Aktív nemterminálisok meghatározása

Legyen most X G egy aktív nemterminálisa. Ekkor létezik G-beli $X = w_n \Rightarrow w_{n-1} \Rightarrow \cdots \Rightarrow w_1 \Rightarrow w$ levezetés ($n \geq 1$), ahol $w \in T^*$.

Állítás: $w_i \in (A_i \cup T)^*$ $(1 \leq i \leq n)$.

i-re vonatkozó teljes indukcióval bizonyítjuk.

$i = 1$-re: az utolsó lépés egy A_1-beli nemterminálist ír át.

Tegyük fel, hogy i-re igaz az állítás. A $w_{i+1} \Rightarrow w_i$ levezetési lépésben átírt nemterminális $w_i \in (A_i \cup T)^*$ egy részszavára íródik át, így ez a részszó is $(A_i \cup T)^*$-beli, vagyis az átírt nemterminális A_{i+1}-beli és így $w_{i+1} \in (A_{i+1} \cup T)^*$.

Tehát, $X = w_n \in (A_n \cup T)^*$, és így $X \in A_n \subseteq A$.

Tehát minden aktív nemterminális A-beli. Ezzel a tételt beláttuk.

Elérhető nemterminálisok meghatározása

Legyen $G = \langle N, T, P, S \rangle$ egy a környezetfüggetlen grammatika

$$R_1 := \{S\},$$

$$R_{i+1} := R_i \cup \{Y \in N \,|\, X \to uYw \in P, X \in R_i, u, w \in (N \cup T)^*\},$$

$$i = 1, 2, \dots$$

Nyilván $R_1 \subseteq R_2 \subseteq \cdots \subseteq N$.

Ebből következik, hogy $\exists k \in \mathbb{N}$, amelyre $R_k = R_{k+1}$. Ekkor $\forall \ell \geq k$ esetén $R_\ell = R_k$.

$R := R_k$.

Tétel

A fenti módon előállított R halmaz éppen a G grammatika elérhető nemterminálisainak halmaza.

Elérhető nemterminálisok meghatározása

Bizonyítás: i-re vonatkozó teljes indukcióval könnyen látható, hogy az R_i-beli nemterminálisok elérhetőek.

$i = 1$-re nyilvánvaló.

Tegyük fel, hogy i-re igaz, és legyen $Y \in R_{i+1}$.

Ekkor létezik $X \in R_i$, $u, w \in (N \cup T)^*$, hogy $p = X \to uYw \in P$. Az indukciós feltevés szerint létezik $v_1, v_2 \in (N \cup T)^*$, melyre $S \Rightarrow^* v_1 X v_2$. Ekkor p alkalmazásával: $S \Rightarrow^* v_1 X v_2 \Rightarrow v_1 u Y w v_2$, tehát Y elérhető.

Elérhető nemterminálisok meghatározása

Másrészt legyen $S = w_1 \Rightarrow \cdots \Rightarrow w_n = uZw$ egy levezetés ($n \geq 1$), ahol $w_i, u, w \in (N \cup T)^*$ $(1 \leq i \leq n)$ és $Z \in N$.

i-re vonatkozó teljes indukcióval látható, hogy w_i nemterminálisai R_i-beliek.

$i = 1$-re nyilvánvaló.

Tegyük fel, hogy i-re igaz és Y w_{i+1}-beli nemterminális. Ha Y w_i-ben is előfordul, akkor az indukciós feltevés miatt $Y \in R_i \subseteq R_{i+1}$. Ha Y nem fordul elő w_i-ben, akkor a $w_i \Rightarrow w_{i+1}$ levezetés során egy $X \to uYw \in P$ szabály került alkalmazásra, ahol $u, w \in (N \cup T)^*$. Az indukciós feltevés miatt $X \in R_i$ és így $Y \in R_{i+1}$.

Tehát $Z \in R_n \subseteq R$, ezzel a tételt beláttuk.

Redukált grammatika

Tétel

Minden környezetfüggetlen G grammatikához létezik vele ekvivalens redukált G' környezetfüggetlen grammatika.

Bizonyítás: Hagyjuk el a G grammatika inaktív nemterminálisait és az őket tartalmazó valamennyi szabályt (G''), majd hagyjuk el az így kapott grammatika nem elérhető nemterminálisait és az ezeket tartalmazó szabályokat (G').

G' nyilván ekvivalens az eredetivel és minden nemterminálisa elérhető. Legyen X G' egy tetszőleges nemterminálisa. G''-ben X aktív volt, azaz levezethető volt belőle G''-ben terminális szó. Mivel X elérhető G'-ben ezért ezen levezetésben előforduló nemterminálisok elérhetőek G'-ben, így X G'-ben is aktív.

Tehát G' G-vel ekvivalens redukált grammatika.

Megjegyzés: Fontos a sorrend, először az inaktívakat majd a nem elérhetőeket hagyjuk el.

Redukált grammatika

Mivel minden 3-as típusú grammatika egyben 2-es típusú is és a fenti bizonyításokban a szabályok közül csak elhagytunk néhányat azonnal adódik a következő:

Következmény

Minden reguláris G grammatikához meg tudunk konstruálni egy vele ekvivalens redukált G' reguláris grammatikát.

Redukált grammatika

Példa

$S \to A \mid bBD$
$A \to AB \mid AC$
$B \to \varepsilon \mid a \mid SS$
$C \to AS \mid aba$
$D \to BB$

Aktívak: $A_1 = \{B, C\}, A_2 = \{B, C, D\}, A_3 = A_4 = \{B, C, D, S\}$, az A-t tartalmazó szabályok elhagyhatók.

$S \to bBD$
$B \to \varepsilon \mid a \mid SS$
$C \to aba$
$D \to BB$

Elérhetők: $R_1 = \{S\}, R_2 = R_3 = \{S, B, D\}$, C elhagyható.

$S \to bBD$
$B \to \varepsilon \mid a \mid SS$
$D \to BB$

Az eredmény az eredetivel ekvivalens és redukált.

Chomsky normálforma

Definíció

Egy $G = \langle N, T, P, S \rangle$ grammatika **Chomsky normálformájú**, ha minden P-beli szabály alakja a következők valamelyike:

- $S \rightarrow \varepsilon$
- $A \rightarrow BC$, $A, B, C \in N$; továbbá $B, C \neq S$, ha $S \rightarrow \varepsilon \in P$.
- $A \rightarrow a$, $A \in N, a \in T$.

Tétel

Minden G környezetfüggetlen grammatikához létezik vele ekvivalens Chomsky normálformájú grammatika.

Bizonyítás: Több lépésben átalakítjuk G-t a megfelelő alakra ügyelve arra, hogy az egyes átalakítási lépések után kapott grammatika G-vel ekvivalens legyen.

Chomsky normálforma

1. lépés Álterminálisok bevezetése

Minden $a \in T$ terminálisra végezzük el a következőt. Legyen $\bar{a}$ egy új nemterminális szimbólum és helyettesítsük minden P-beli szabály jobboldalán a előfordulásait $\bar{a}$-val. Adjuk hozzá ezen kívül P-hez a $\bar{a} \to a$ új szabályt.

Az így kapott grammatika az eredetivel ekvivalens (ezt már korábban meggondoltuk, a zártsági tétel bizonyításánál).

Kivétel: Amennyiben $A \to a \in P$, akkor a-t nem szükséges átírni, hiszen ez a szabály már kellő alakú.

Chomsky normálforma

2. lépés Hosszredukció

Ekkor minden $X \to Y_1 Y_2 \cdots Y_k$, $k \geq 3$ alakú szabályt helyettesítünk egy

$$\{X \to Y_1 Z_1, Z_1 \to Y_2 Z_2, \ldots, Z_{k-2} \to Y_{k-1} Y_k\}$$

szabályhalmazzal, ahol $Z_1, \ldots, Z_{k-2}$ új, a szabályhoz bevezetett nemterminálisok.

Nyilván minden $X \to Y_1 Y_2 \cdots Y_k$ szabályalkalmazás helyettesíthető a fenti szabályokkal. Másrészt minden, valamelyik Z_i-t tartalmazó levezetés az átalakított grammatikában $Z_1, \ldots, Z_{k-2}$ sorrendben minden Z_i-t tartalmaz, és meggondolható, hogy amennyiben a Z_i-t behozó lépést nem közvetlenül követi a Z_i-t átíró lépés, akkor ez a Z_i-t átíró lépés a levezetésben előrehozható. Tehát feltehető, hogy a Z_i-k a fenti sorrendben, közvetlenül egymást követő lépésekben alakulnak át, vagyis $X \to Y_1 Y_2 \cdots Y_k$ alkalmazását szimulálják.

Chomsky normálforma

3. lépés ε-mentesítés

Definiáljuk az $U_i \subseteq N$ halmazokat a következőképpen:

$$U_1 := \{X \mid X \to \varepsilon \in P\},$$

$$U_{i+1} := U_i \cup \{X \mid X \to u \in P \text{ és } u \in U_i^*\}. \quad i \geq 1$$

Nyilvánvaló, hogy az U_i sorozat, $i = 1, 2, \ldots$ a tartalmazásra nézve hierarchiát alkot és van olyan k index, hogy $U_k = U_{k+1}$ és így $U_\ell = U_k$ minden $\ell \geq k$-ra.

$U := U_k$.

Ekkor az aktív nemterminálisok halmazának algoritmikus előállításánál látotthoz hasonló gondolatmenettel meggondolható, hogy $X \Rightarrow_G^* \varepsilon$ akkor és csak akkor, ha $X \in U$.

Ebből rögtön következik, hogy $\varepsilon \in L(G)$ akkor és csak akkor, ha $S \in U$.

Chomsky normálforma

Hagyjuk el P-ből az összes $X \to \varepsilon$ alakú ún. **ε-szabályt** és adjuk hozzá a következőket

- $A \to BC \in P, B, C \in U$ esetén az $A \to B$ és $A \to C$ szabályokat
- $A \to BC \in P, B \in U, C \notin U$ esetén az $A \to C$ szabályt
- $A \to BC \in P, B \notin U, C \in U$ esetén az $A \to B$ szabályt

Legyen G_1 az így kapott grammatika. Ekkor $L(G_1) \subseteq L(G) \setminus \{\varepsilon\}$, hiszen minden új szabály alkalmazása megfelel egy régi szabály alkalmazásának amelyet egy $B \Rightarrow_G^* \varepsilon$ vagy $C \Rightarrow_G^* \varepsilon$ levezetés alkalmazásával kombinálunk.

Másrészt $L(G) \setminus \{\varepsilon\} \subseteq L(G_1)$. Ugyanis, ha $S \Rightarrow_G^* u$ és $u \neq \varepsilon$, akkor $S \Rightarrow_{G'}^* u$, hiszen az $X \to \varepsilon$ típusú szabályok alkalmazása elkerülhető egy megfelelő új szabály alkalmazásával.

Chomsky normálforma

a) eset $S \notin U$

$S \notin U$ esetén G_1 ekvivalens G-vel, és vegyük észre, hogy a KES szabály is teljesül, mivel a grammatikának egyáltalán nincs ε-szabálya.

b) eset $S \in U$

$S \in U$ esetén legyen G_1' a G_1 grammatika következő módosítása: adjuk hozzá G_1 nemterminálisainak halmazához egy új, S' nemterminálist, és legyen S' a G_1' kezdőszimbóluma. G_1' szabályrendszere tartalmazza G_1 összes szabályát és ezen felül még az $S' \to \varepsilon$ és $S' \to S$ szabályokat.

Ekkor $L(G_1') = L(G)$, egyetlen ε-szabálya az $S' \to \varepsilon$ szabály, S' nem fordul elő szabály jobboldalán, így G_1'-re teljesül a KES is.

Kivétel: Amennyiben $S \in U$, de S nem fordul elő szabály jobboldalán, akkor nem szükséges egy új S' kezdőszimbólum hozzáadása, elegendő az $S \to \varepsilon$ szabályt hozzáadni G_1 szabályaihoz.

Chomsky normálforma

4. lépés Láncmentesítés

Már csak az $X \to Y$ alakú ($X, Y \in N$), ú.n. **láncszabályokat** kell eliminálni. Jelölje P_0 a P-beli láncszabályok halmazát.

Minden $A \in N$ esetén legyen $H(A) := \{B \in N \mid A \Rightarrow^* B\}$.

A G-vel ekvivalens, láncmentes G' grammatika szabályai:

$P' := \{A \to w \mid \exists B \in H(A), \text{ amelyre } B \to w \in P\} \setminus P_0$.

$L(G') \subseteq L(G)$, hiszen egy $A \to w$ új szabály alkalmazása megfelel az $A \Rightarrow^* B$ láncszabályok és a $B \to w$ eredeti szabály alkalmazásának ($B \in H(A)$).

Másrészt $L(G) \subseteq L(G')$, hiszen az új szabályokkal a láncszabályok alkalmazása elkerülhető.

Chomsky normálforma

$H(A)$ algoritmikus előállításához definiáljuk rekurzívan a $H_i(A)$ $(i \geq 0)$ halmazokat az alábbiak szerint:

$H_0(A) := \{A\}$

$H_{i+1}(A) := H_i(A) \cup \{B \in N \mid \exists C \in H_i(A)$ amelyre $C \to B \in P\}$.

$H_0(A) \subseteq H_1(A) \subseteq \cdots \subseteq H_k(A) \subseteq \cdots \subseteq N$

$k := \min\{0 \leq i \leq n-1 \mid H_i(A) = H_{i+1}(A)\}$.

$H(A) := H_k(A)$.

Ekkor az elérhető nemterminálisok halmazának algoritmikus előállításánál látotthoz hasonló gondolatmenettel meggondolható, hogy $H_k(A)$ valóban az A-ból láncszabályokkal elérhető nemterminálisok halmaza. $\square$

Chomsky normálforma – példa

Példa: Hozzuk az alábbi grammatikát Chomsky normálformára! (a nagybetűk a nemterminálisok, S a kezdőszimólum)

$S \longrightarrow AB$
$A \longrightarrow aAa \mid C$
$B \longrightarrow bBb \mid C$
$C \longrightarrow Cabc \mid b \mid \varepsilon$

Megoldás:

1. lépés (Álterminálisok bevezetése):

$S \longrightarrow AB$
$A \longrightarrow DAD \mid C$
$B \longrightarrow EBE \mid C$
$C \longrightarrow CDEF \mid b \mid \varepsilon$
$D \longrightarrow a$
$E \longrightarrow b$
$F \longrightarrow c$

Chomsky normálforma – példa

2. lépés (Hosszredukció):

$$
\begin{aligned}
S &\rightarrow AB \\
A &\rightarrow DAD \mid C \\
B &\rightarrow EBE \mid C \\
C &\rightarrow CDEF \mid b \mid \varepsilon \\
D &\rightarrow a \\
E &\rightarrow b \\
F &\rightarrow c
\end{aligned}
\quad \Longrightarrow \quad
\begin{aligned}
S &\rightarrow AB \\
A &\rightarrow DZ_1 \mid C \\
B &\rightarrow EZ_2 \mid C \\
C &\rightarrow CZ_3 \mid b \mid \varepsilon \\
D &\rightarrow a \\
E &\rightarrow b \\
F &\rightarrow c \\
Z_1 &\rightarrow AD \\
Z_2 &\rightarrow BE \\
Z_3 &\rightarrow DZ_4 \\
Z_4 &\rightarrow EF
\end{aligned}
$$

Chomsky normálforma – példa

3. lépés (ε-mentesítés):

$U_1 = \{C\}$, $U_2 = \{C, A, B\}$, $U_3 = U_4 = \{S, A, B, C\} = U$.

$$
\begin{array}{lll}
S \to AB & & S \to AB \mid A \mid B \mid \varepsilon \\
A \to DZ_1 \mid C & & A \to DZ_1 \mid C \\
B \to EZ_2 \mid C & & B \to EZ_2 \mid C \\
C \to CZ_3 \mid b \mid \varepsilon & & C \to CZ_3 \mid b \mid Z_3 \\
D \to a & & D \to a \\
E \to b & \Longrightarrow & E \to b \\
F \to c & & F \to c \\
Z_1 \to AD & & Z_1 \to AD \mid D \\
Z_2 \to BE & & Z_2 \to BE \mid E \\
Z_3 \to DZ_4 & & Z_3 \to DZ_4 \\
Z_4 \to EF & & Z_4 \to EF
\end{array}
$$

Megjegyzés: $S \in U$, ezért a b) esetet alkalmaztuk, de S nem szerepel szabály jobboldalán, így most ebben a speciális esetben nincs szükség új kezdőszimbólum bevezetésére.

Chomsky normálforma – példa

4. lépés (Láncmentesítés):
$H_0(S) = \{S\}, H_1(S) = \{S, A, B\}, H_2(S) = \{S, A, B, C\}, H_3(S) = \{S, A, B, C, Z_3\} = H(S)$

Hasonlóan $H(A) = \{A, C, Z_3\}, H(B) = \{B, C, Z_3\}$, $H(C) = \{C, Z_3\}$, $H(Z_1) = \{D, Z_1\}$, $H(Z_2) = \{E, Z_2\}$, a többi csak önmagát tartalmazza.

$$
\begin{array}{lcl}
S \to AB \mid A \mid B \mid \varepsilon & & S \to AB \mid DZ_1 \mid CZ_3 \mid b \mid DZ_4 \mid EZ_2 \mid \varepsilon \\
A \to DZ_1 \mid C & & A \to DZ_1 \mid CZ_3 \mid b \mid DZ_4 \\
B \to EZ_2 \mid C & & B \to EZ_2 \mid CZ_3 \mid b \mid DZ_4 \\
C \to CZ_3 \mid b \mid Z_3 & & C \to CZ_3 \mid b \mid DZ_4 \\
D \to a & & D \to a \\
E \to b & \Longrightarrow & E \to b \\
F \to c & & F \to c \\
Z_1 \to AD \mid D & & Z_1 \to AD \mid a \\
Z_2 \to BE \mid E & & Z_2 \to BE \mid b \\
Z_3 \to DZ_4 & & Z_3 \to DZ_4 \\
Z_4 \to EF & & Z_4 \to EF
\end{array}
$$

Chomsky normálforma – az algoritmus hatékonysága

Legyen $G = \langle N, T, P, S \rangle$ egy környezetfüggetlen grammatika. Ha $p : X \to Y_1 \cdots Y_m \in P$, akkor legyen $|p| := m + 1$, a p szabály bal- és jobboldalának összhossza. Jelölje $|G| := \sum_{p \in P} |p|$ a G grammatika **méretét**.

A Chomsky normálformára hozás hatékonysága:
A fenti algoritmus nagyságrendileg $|G|^2$ lépésben előállít egy nagyságrendileg legfeljebb $|G|^2$ méretű G-vel ekvivalens Chomsky normálformájú grammatikát.

Megjegyzés: A méretnövekedés a láncmentesítés kivételével lineáris.

A környezetfüggetlen grammatikák szóproblémája

Állítás: Legyen $G = \langle T, N, P, S \rangle$ környezetfüggetlen grammatika és egy $u \in T^*$ egy szó. Eldönthető, hogy $u \overset{?}{\in} L(G)$.

Bizonyítás: Az előző algoritmus alapján algoritmikusan előállítható egy G-vel ekvivalens Chomsky normálformájú grammatika. Legyen $n = |u|$.

$n = 0$ esetén $u \in L(G) \;\Leftrightarrow\; S \rightarrow \varepsilon \in P$.

Teljes indukcióval könnyen látható, hogy $n \geq 1$ esetén u levezetése $n - 1$ darab „$A \rightarrow BC$" típusú és n darab „$A \rightarrow a$" típusú szabály alkalmazásából áll. Mivel adott grammatika és n esetén véges sok ilyen levezetés van ezek megvizsgálása után el tudjuk dönteni, hogy $u \overset{?}{\in} L(G)$.

A környezetfüggetlen grammatikák szóproblémája

Az előző algoritmus alapján csak exponenciális felső becslést adhatunk a szóprobléma eldöntésének hatékonyságára.

u levezetéseinek keresését azonban hatékonyabban is megszervezhetjük egy „bottom-up” ún. „dinamikus programozás” segítségével, az így kapott algoritmus már polinomiális (azaz hatékony) lesz.

John Cocke, Daniel Younger és Tadao Kasami algoritmusa (CYK algoritmus) egy Chomsky normálformában adott grammatika esetében $|G| \cdot n^3$ nagyságrendű lépésben eldönti a szóproblémát.

Tetszőleges környezetfüggetlen grammatika esetén tehát a Chomsky normálformára hozással kombinálva a CYK algoritmus hatékonyságára $|G|^2 \cdot n^3$ nagyságrendű felső becslés adható.

CYK algoritmus 2-es típusú szóproblémára

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

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

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

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

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

A CYK algoritmus helyessége

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 $X \Rightarrow YZ \Rightarrow^* t_i \cdots t_h Z \Rightarrow^* t_i \cdots t_j$.

CYK algoritmus

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:

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

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

CYK algoritmus

$$S \longrightarrow AB \mid BC$$
$$A \longrightarrow XA \mid a$$
$$X \longrightarrow a$$
$$C \longrightarrow YC \mid c$$
$$Y \longrightarrow c$$
$$B \longrightarrow UV \mid VW \mid XS$$
$$U \longrightarrow XX$$
$$W \longrightarrow YY \mid XS$$
$$V \longrightarrow ZZ$$
$$Z \longrightarrow b$$

				$\{Y, C\}$	
a	a	b	b	c	c

CYK algoritmus

$S \rightarrow AB \mid BC$

$A \rightarrow XA \mid a$

$X \rightarrow a$

$C \rightarrow YC \mid c$

$Y \rightarrow c$

$B \rightarrow UV \mid VW \mid XS$

$U \rightarrow XX$

$W \rightarrow YY \mid XS$

$V \rightarrow ZZ$

$Z \rightarrow b$

$\{A, U\}$					
$\{A, X\}$	$\{A, X\}$	$\{Z\}$	$\{Z\}$	$\{Y, C\}$	$\{Y, C\}$
a	a	b	b	c	c

$$\{A, X\}\{A, X\} = \{AA, AX, XA, XX\}$$

CYK algoritmus

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

	$\{\ \}$				
$\{A, U\}$	$\{\ \}$	$\{V\}$	$\{\ \}$	$\{C, W\}$	
$\{A, X\}$	$\{A, X\}$	$\{Z\}$	$\{Z\}$	$\{Y, C\}$	$\{Y, C\}$
a	a	b	b	c	c

$\{A, X\}\{\ \} = \{\ \}$ $\quad$ $\{A, U\}\{Z\} = \{AZ, UZ\}$

CYK algoritmus

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

		$\{B\}$			
$\{\,\}$	$\{\,\}$	$\{\,\}$	$\{\,\}$		
$\{A, U\}$	$\{\,\}$	$\{V\}$	$\{\,\}$	$\{C, W\}$	
$\{A, X\}$	$\{A, X\}$	$\{Z\}$	$\{Z\}$	$\{Y, C\}$	$\{Y, C\}$
a	a	b	b	c	c

$\{Z\}\{\,\} = \{\,\}$ $\{V\}\{C, W\} = \{VC, VW\}$ $\{\,\}\{Y, C\} = \{\,\}$

CYK algoritmus

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

$\{S\}$					
$\{B\}$	$\{\ \}$	$\{B\}$			
$\{\ \}$	$\{\ \}$	$\{\ \}$	$\{\ \}$		
$\{A, U\}$	$\{\ \}$	$\{V\}$	$\{\ \}$	$\{C, W\}$	
$\{A, X\}$	$\{A, X\}$	$\{Z\}$	$\{Z\}$	$\{Y, C\}$	$\{Y, C\}$
a	a	b	b	c	c

$\{\ AB, XB\ \}$ $\{\ \}$ $\{\ \}$ $\{\ \}$

CYK algoritmus

$S \rightarrow AB \mid BC$

$A \rightarrow XA \mid a$

$X \rightarrow a$

$C \rightarrow YC \mid c$

$Y \rightarrow c$

$B \rightarrow UV \mid VW \mid XS$

$U \rightarrow XX$

$W \rightarrow YY \mid XS$

$V \rightarrow ZZ$

$Z \rightarrow b$

$\{S, B, W\}$					
$\{S\}$	$\{S\}$				
$\{B\}$	$\{\}$	$\{B\}$			
$\{\}$	$\{\}$	$\{\}$	$\{\}$		
$\{A, U\}$	$\{\}$	$\{V\}$	$\{\}$	$\{C, W\}$	
$\{A, X\}$	$\{A, X\}$	$\{Z\}$	$\{Z\}$	$\{Y, C\}$	$\{Y, C\}$
a	a	b	b	c	c

$\{AS, XS\}$ $\{AB, UB\}$ $\{\}$ $\{BC, BW\}$ $\{SY, SC\}$

CYK algoritmus

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

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

Mivel $S \in H_{1,6}$, ezért $aabbcc \in L(G)$.
Visszafejthető például a következő levezetés:
$S \Rightarrow AB \Rightarrow XAB \Rightarrow XAVW \Rightarrow XAZZW \Rightarrow XAZZYY \Rightarrow^* aabbcc$.