Redukált gramatika, Chomsky NF

Ez az anyag 2. típusú grammatikákról szól!!!

A) Elméleti háttér

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

A környezetfüggetlen grammatika egy nemterminálisát **nem elérhetőnek** nevezzük, ha nem fordul elő egyetlen olyan mondatformában sem, amely a kezdőszimbólumból levezethető; egyébként **elérhetőnek** mond- juk.

Egy nemterminálist **nem hasznosnak** mondunk, ha vagy inaktív, vagy nem elérhető, vagy mindkét tulajdonság teljesül esetében. Egy nem- terminálist **hasznosnak** nevezünk, ha aktív és elérhető.

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

Tétel: Minden környezetfüggetlen grammatikához meg tudunk konstruálni egy vele ekvivalens redukált környezetfüggetlen grammatikát.

Aktív nemterminálisok:

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

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

$A_1 \subseteq A_2 \subseteq \cdots \subseteq N$

Ha $A_k = A_{k+1}$ akkor megkaptuk az aktív nemterminálisokat.

Az inaktívakat tartalmazó szabályokat elhagyjuk.

Elérhető nemterminálisok:

$R_1 = \{S\}$,

$R_{i+1} = R_i \cup \{Y \mid X \to uYw \in P, X \in R_i, u, w \in (N \cup T)^*\}$, $i = 1, 2, \ldots$

$R_1 \subseteq R_2 \subseteq \cdots \subseteq N$

Ha $R_k = R_{k+1}$ akkor megkaptuk az aktív nemterminálisokat.

A nem elérhetőket elhagyjuk.
Chomsky normálforma
Lehetséges szabályalakok:

$S \to \varepsilon$
$A \to BC$, $A, B, C \in N$; továbbá $B, C \neq S$, ha van $S \to \varepsilon$ szabály.
$A \to a$, $A \in N, a \in T$.

Tétel: Minden környezetfüggetlen grammatikához meg tudunk konstruálni egy vele ekvivalens Chomsky NF-jú grammatikát.

Lépések:

1. Ha van S a jobboldalon, vezessük be $S_0 \to S$ új szabályt és S_0 legyen az új kezdőszimbólum.

2. Álterminálisok bevezetése: Minden, valamely szabály jobboldalán álló, de nem önmagában álló a terminálist cseréljünk le egy neki megfelelő X_a nemterminálisra és adjuk hozzá az $X_a \to a$ szabályt a grammatikához.

3. Hosszredukció: példa: $X \to ABCD$ helyett $X \to AZ_1$, $Z_1 \to BZ_2$, $Z_2 \to CD$, ahol Z_1, Z_2 új nemterminálisok.

4. ε-mentesítés

Először konstruálunk egy $H \subseteq N$ segédhalmazt, melynek pontosan azok a nemterminálisok lesznek az elemei, melyekből (egy vagy több lépésben) levezethető ε. Ehhez ismét rekurzívan definiálunk H_i halmazokat. A H_i halmaz a H_{i-1} halmaz bővítése azon nemterminálisokkal, amelyekből közvetlenül levezethető H_{i-1}^*-beli szó. A kiindulási halmaz H_1, azon nemterminálisok halmaza, melyekből közvetlenül levezethető ε.

Ezután képezzük az összes olyan szabályt, melynek jobboldala valamely eredeti szabály jobboldalából tetszőlegesen kiválasztott H-beli nemterminálisok elhagyásával kapható (beleértve azt az esetet is, ha nem hagyunk el semmit), a baloldal pedig az eredeti baloldal. Az ε-szabályokat elhagyjuk.

Végül ha a kezdőszimbólum H-beli, akkor adjunk a grammatikához egy rá vonatkozó ε-szabályt (ha eddig nem volt).

5. Láncmentesítés

Már csak az $A \to B$ $(A, B \in N)$ alakú szabályok nem kellő alakúak. Ezekkel ugyanúgy bánunk el, mint a 3. típus esetén.

Meghatározzuk minden nyelvtani jelhez (az adott nyelvtani jelet magát is beleértve) azon nemterminálisok $H(X)$ halmazát, melyek mint 1 hosszúságú szó (közvetlenül, vagy közvetetten) levezethetők belőle

Minden $Y \in H(X)$ nyelvtani jelhez vesszük azon szabályokat, amelyeknek baloldalán X, jobboldalán pedig egy Y-ra vonatkozó eredeti szabály jobboldala áll, kivéve ha ez a jobboldali mondatforma csupán egyetlen nemterminálisból áll. (Azaz, a láncszabályokat elhagyjuk.)

A $H(X)$ halmazok a szokásos rekurzív közelítéssel állíthatók elő. $H_0(X) := \{X\}$, majd ezt sorra bővítjük a lépésenként felfedezett újabb láncszabálysorozatok végpontjaival.

B) Mintapéldák:

1. példa, redukálás

$S \to A \,|\, bBD$
$A \to AB \,|\, A$
$B \to \varepsilon \,|\, a \,|\, SS$
$C \to AS \,|\, a$
$D \to BB$

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

$S \to bBD$
$B \to \varepsilon \,|\, a \,|\, SS$
$C \to a$
$D \to BB$

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

$S \to bBD$
$B \to \varepsilon \,|\, a \,|\, SS$
$D \to BB$

Az eredmény az eredetivel ekvivalens és redukált.

2. példa: Hozzuk Chomsky NF-ra!

$S \to AB$
$A \to aAa \,|\, C$
$B \to bBb \,|\, C$
$C \to Cabc \,|\, b \,|\, \varepsilon$

Megoldás:

1. lépés: Nincs S a jobboldalon, maradhat S a kezdőszimbólum.

2. lépés:

$S \to AB$
$A \to DAD \,|\, C$
$B \to EBE \,|\, C$
$C \to CDEF \,|\, b \,|\, \varepsilon$
$D \to a$
$E \to b$
$F \to c$

3. lépés:

$S \to AB$
$A \to DZ_1 \,|\, C$
$B \to EZ_2 \,|\, C$
$C \to CZ_3 \,|\, b \,|\, \varepsilon$
$D \to a$
$E \to b$
$F \to c$
$Z_1 \to AD$
$Z_2 \to BE$
$Z_3 \to DZ_4$
$Z_4 \to EF$

4. lépés:

$H_0 = \{C\}$, $H_1 = \{C, A, B\}, H_2 = H_3 = \{S, A, B, C\} = H.$

$S \to AB \,|\, A \,|\, B \,|\, \varepsilon$
$A \to DZ_1 \,|\, C$
$B \to EZ_2 \,|\, C$
$C \to CZ_3 \,|\, b \,|\, Z_3$
$D \to a$
$E \to b$
$F \to c$
$Z_1 \to AD \,|\, D$
$Z_2 \to BE \,|\, E$
$Z_3 \to DZ_4$
$Z_4 \to EF$

5. lépés:

$H_0(S) = \{S\}, H_1(S) = \{S, A, B\}, H_2(S) = \{S, A, B, C\}, H_3(S) = \{S, A, B, C, Z_3\} = H(S)$

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.

$S \to AB \,|\, DZ_1 \,|\, CZ_3 \,|\, b \,|\, DZ_4 \,|\, EZ_2 \,|\, \varepsilon$
$A \to DZ_1 \,|\, CZ_3 \,|\, b \,|\, DZ_4$
$B \to EZ_2 \,|\, CZ_3 \,|\, b \,|\, DZ_4$
$C \to CZ_3 \,|\, b \,|\, DZ_4$
$D \to a$
$E \to b$
$F \to c$
$Z_1 \to AD \,|\, a$
$Z_2 \to BE \,|\, b$
$Z_3 \to DZ_4$
$Z_4 \to EF$

C) Gyakorló feladatok:

1. feladat: Határozzuk meg az alábbi $G = (\{S, A, B, C, D\}, \{a, b\}, P, S)$ környezetfüggetlen grammatika **aktív** nemterminálisait a tanult algoritmus alapján! A P szabályrendszer:

$S \to BaB \,|\, DaD$
$A \to aBC \,|\, AS$
$B \to SbDD \,|\, A$
$C \to DSC \,|\, aA$
$D \to ab \,|\, AC$

2. feladat: Hozzuk *Chomsky normálformára* a tanult algoritmus alapján az alábbi nyelvtant (S a kezdőszimbólum, S, A, B, C, D a nemterminálisok, a, b a terminálisok):

$S \to AB \,|\, AC$
$A \to aba \,|\, DS$
$B \to DCC \,|\, aS$
$C \to bbD \,|\, \varepsilon$
$D \to SS \,|\, \varepsilon$