Kontekstivabad grammatikad

Meeldetuletuseks: KV(kontekstivaba) grammatika on nelik

- N – mitteterminaalide tähestik
- $\sum$ - terminaalide tähestik
- P – reeglite (produktsioonide) hulk
- S – algsümbol

KV grammatika genereerib keele, mis koosneb kõigist terminaalsetest sõnadest, mida saab reegleid rakendades algsümbolist tuletada:

$L(G) = \{w \in \sum^* \mid S =>^* w \}$

Sümbol => tähendab üht tuletussammu, st ühe korra ühe reegli rakendamist, $=>^*$ aga mingi arv kordi (null või enam) reeglite rakendamist.

KV grammatika reeglite vasakul poolel on alati täpselt üks mitteterminaal ja reegli rakendamine tähendab seda, et vasakul poolel olev sümbol asendatakse sellega, mis on kirjas reegi paremal poolel. Püstkriips eraldab erinevaid reegi alternatiive, nt $S \rightarrow XSX|R$ tähendab, et võime reegli rakendamisel asendada S kas XSX või R-ga.

Selles osas kasutame keelte kirjeldamisel ka astmeid, st. kui a^* tähendab 0 või enamat järjestikust a-d, siis a^n tähendab n järjestikust sümbolit a. Näiteks a^4 on seega $aaaa$.

Ülesanded.

1. Olgu antud grammatika ($\Sigma = \{a,b\}$, algsümbol on S):

$$S \rightarrow XSX \mid R$$

$$R \rightarrow aTb \mid bTa$$

$$T \rightarrow XTX \mid X \mid \varepsilon$$

$$X \rightarrow a|b$$

Kas kehtib, et:

$T \Rightarrow aba$ Ei, sest see oleks õige ainult siis, kui reeglite hulgas oleks reegel $T \rightarrow aba$ (tuletuskäik pikkusega 1).

$T \Rightarrow^* aba$ Jah, sest saame tuletada $T \Rightarrow XTX \Rightarrow aTX \Rightarrow aXX \Rightarrow$

$\Rightarrow abX \Rightarrow aba$, (mingi arv kordi reeglite rakendamine). Samas ei saaks me aga tuletada $S \Rightarrow aba$, sest sümbolist S vabanemiseks peaksime kasutama reeglit $S \rightarrow R$ ja siis kas $R \rightarrow aTb$ või $R \rightarrow bTa$, millest ei saa kuidagi tuletada sõna *aba.*

$T \Rightarrow T$ Ei, sest pole reeglit $T \rightarrow T$.

$T \Rightarrow^* T$ Jah, sest * sisaldab ka null korda rakendamist.

$XXX \Rightarrow^* aba$ Jah, tuletame $XXX \Rightarrow aXX \Rightarrow abX \Rightarrow aba$.

$X \Rightarrow^* aba$ Ei, sest ühest X-st saame ainult kas ühe a või ühe b.

$T \Rightarrow^* XX$ Jah, $T \Rightarrow XTX \Rightarrow XX$ (kasutades reeglit $T \rightarrow \varepsilon$)

$S \Rightarrow^* \varepsilon$ Ei, sest sümbolist S vabanemiseks tuleks kasutada reeglit $S \rightarrow R$ ja siis kas $R \rightarrow aTb$ või $R \rightarrow bTa$, millest ei saa kuidagi tuletada tühja sõna.

Kirjelda, millise keele antud grammatika genereerib.

Esimese reegliga saame tuletada selliseid sõnu, kus paremal ja vasakul pool sümbolit S on mistahes arv X-e, oluline on aga see, et mõlemal pool on neid võrdselt. Kui on vaja genereerida igassuguse pikkusega sõnu, tuleb kasutada rekursiivseid reegleid – ehk siis vasakul poolel olev sümbol ilmub uuesti paremal poolel. Ja oluline on ka see, et oleks võimalus rekursiivne pöördumine lõpetada – ehk siis reeglil on olemas ka mitterekursiivne alternatiiv. Igast sümbolist X saab tuletada ühe a või b, seega tuleb lõpuks mõlemale poole sümbolit S samapikkune, aga muidu suvaline järjend tähtedest a ja b. Kuna tuletatavasse keelde saavad kuuluda ainult terminaalidest koosnevad sõnad, siis tahame me mingil hetkel sümbolist S vabaneda ja kasutame teist reeglit, $S \rightarrow R$. Seejärel peame kasutama kas reeglit kolm või neli, ja oleme olukorras, kus vahetult keskmise sümboli T kõrval on ühel pool a ja teisel pool b. Edasi võime lisada jälle mõlemale poole keskmist sümbolit X reegliga $T \rightarrow XTX$, saades jälle võrdsed, kuid muidu suvalise pikkusega järjendid sümbolitest X. Viimaks võime lisada veel ühe üksiku sümboli reegliga $T \rightarrow X$ või tühja sõna ε, ja viimaks asendada kõik sümbolid X kas a või b-ga. Iga X puhul on vaba valik, kas asendame selle a või b-ga. Seega kuuluvad meie genereeritud keelde kõikvõimalikud järjendid sümbolitest a ja

b, mis ei ole algusest ja lõpust lugedes üks ja sama sõna (leidub vähemalt üks positsioon, kus ühelt poolt lugedes on a ja teiselt poolt lugedes on b).

2. Leida KV grammatika, mis genereerib keele ($\Sigma = \{0,1\}$)

2.1. $L = \{w \mid w$ sisaldab vähemalt kahte 1$\}$

Kui ma tahan tagada, et kaks ühte oleksid igas sõnas, siis kirjutan selle esimesse reeglisse sisse: $S \to X1X1X$.

Sümbolid X on eelmises reeglis selleks, et saaksin võimalikku kolme positsiooni lisada veel sümboleid (näiteks muidu ei saaks ma genereerida kõiki neid sõnu, kus ongi ainult kaks ühte ja ülejäänud nullid).

Järgmiseks lisan reegli, mis võimaldab mul tekitada igast sümbolist X igasuguse pikkusega sõnu: $X \to 0X \mid 1X \mid 0 \mid 1 \mid \varepsilon$

Samaväärne oleks teha nii: $X \to AX \mid A \mid \varepsilon,\ A \to 0 \mid 1$

2.2. $L = \{w \mid w$ algab ja lõpeb sama sümboliga$\}$

KV grammatika puhul on võrdlemist kõige mõistlikum teha nii, et võrreldavad sõnaosad genereeritakse ühes ja samas reeglis. Seega panen esimeses reeglis kohe sõna alguse ja lõpu paika: $S \to 0X0 \mid 1X1 \mid 0 \mid 1$

Kolmas ja neljas alternatiiv on selleks, kui sõna koosnebki ainult ühest sümbolist – aga ka siis ta ju algab ja lõpeb sama sümboliga.

Nüüd on meil tingimus täidetud ja peame saama genereerida igasuguse pikkusega järjendeid sümbolitest 1 ja 0: $X \to 0X \mid 1X \mid \varepsilon$

2.3. $L = \{w \mid w$ pikkus on paaritu$\}$

Sõna pikkus püsib paarituna, kui lisame ühele sümbolile paarikaupa sümboleid juurde, seega: $S \to XXS \mid X$

Samaväärselt võiks teha: $S \to SXX \mid X$

Või: $S \to XSX \mid X$

Kui oleme pikkuse paika pannud: $X \to 0 \mid 1$

Viimane reegel võimaldab meil kõik mitteterminaalid asendada terminaalidega. Siin tasub panna tähele, et X jaoks meil seekord rekursiivset reeglit pole – muidu läheks pikkus paigast ära ja grammatika hakkaks genereerima valesid sõnu.

2.4. $L = \{w \mid w$ pikkus on paaritu ja keskmine sümbol on $1\}$

Kuna keskmine sümbol peab olema 1, siis tuleb paaritu pikkuse saavutamiseks sümboleid lisada kahele poole keskmist sümbolit: $S \rightarrow XSX \mid 1$

Teine alternatiiv lubab meil rekursiooni lõpetada ja jätta 1 keskmiseks.

Peame vabanema ja kõigist sümbolitest *X*, ja kuna oleme neid paarikaupa juurde lisanud, ei luba me neid enam mujal lisada, ehk siis: $X \rightarrow 0 \mid 1$

3. Leida KV grammatika, mis genereerib keele ($\Sigma = \{a,b\}$):

3.1. $L = \{a^n b^m \mid n > m,\ \ m \geq 0\}$

Seega meie keele sõnades on kõik sümbolid *a* enne sümboleid *b* ning *a*-sid on rohkem kui *b*-sid. Võrdlemiseks genereerin jälle sümboleid koos ühes ja samas reeglis: $S \rightarrow aSb \mid aS \mid a$

Esimene alternatiiv tagab selle, et iga lisanduva *b*-ga lisandub sõnasse ka *a*, teine alternatiiv lubab sümboleid *a* lisada ka üksi, st nende arv saab minna suuremaks *b*-de arvust, ja viimane alternatiiv võimaldab sümbolist *S* lõpuks vabaneda. Vähim sõna, mis sellesse keelde kuulub, on *a*. Aga samuti kuuluvad siia nt *aab, aaaab, aaaaaaabbb* jne. Oluline on jälgida, et ei genereeritaks valesid sõnu. Kui meie reegel oleks hoopis $S \rightarrow abS$, siis tekiksid hoopis sellised sõnad: *aba, abababaaaaba* jne.

3.2. $L = \{w \mid$ sõnas *w* on *a*-sid on rohkem kui *b*-sid$\}$

Erinevalt eelmisest ülesandest võib *a*-de ja *b*-de järjekord nüüd olla igasugune.

Esimese reegliga kindlustame, et vähemalt üks *a* on rohkem: $S \rightarrow TaT$

Miks on *a* kahe mitteterminaali vahel? Sest kui eeldame, et see on alati alguses või lõpus, kitsendame on sõnade valikut. (Seega reeglitest $S \rightarrow aT$ ja $S \rightarrow Ta$ jääks väheks.)

Järgmiseks kindlustame, et koos iga *b*-ga lisandub sõnasse alati ka üks *a*. Samas lubame sümbolit *a* ka üksinda lisada ja ka sümbolist *T* vabaneda:

$T \rightarrow aTb \mid bTa \mid a \mid \varepsilon$

Sellisel moel lisamine aga kitsendab tekkivate sõnade struktuuri, sellepärast on meil vaja veel üht reeglit: $T \rightarrow TT$

Millisel hetkel seda vaja võiks minna? Konkreetne näide: sõna *aabbbabaa*. Kui viimast reeglit poleks, ei saaks seda sõna genereerida (Kontrollige ise järele, miks.)

Märkus. Kui teil on kahtlusi oma grammatika õigsuses, proovige jälle järele erinevate sõnadega, mis keelde kuuluvad – kas neid saab ikka tuletada. Ja teiselt poolt, tuleks ka kontrollida, et ei genereeritaks valesid sõnu – selleks tuleks läbi proovida erinevaid tuletuskäike, et näha, millised sõnad tekivad.

3.3. $L = \{a^i b^j c^k \mid i = j, \quad i, j, k \geq 0\}$ ($\Sigma = \{a,b,c\}$)

Meil tuleb võrrelda sümbolite a ja b arvu, sümboleid c võib aga olla kuitahes palju. Siin jälle mängus tähtede õige järjekord, st. tulemussõnades ei tohi eri sümbolid läbisegi asuda. Teeme oma tuletuse kaheks osaks: $S \rightarrow AC$

Mitteterminaali A kaudu lisame võrdsel arvul ja samas ka õiges paigutuses sümbolid a ja b. $A \rightarrow aAb \mid \varepsilon$

Kuna reegli paremal poolel asub A keskel, tekivad kõik a-d sellest vasakule ja kõik b-d sellest paremale. Teine alternatiiv on tuletuse lõpetamiseks.

Mitteterminaali C kaudu tuletame kõik sümbolid c: $C \rightarrow cC \mid \varepsilon$