Модели вычислений

В.А. Захаров, Р.И. Подловченко

Лекция 7.

1. Формальные грамматики.
2. Иерархия Хомского.
3. Регулярные грамматики
4. Неограниченные грамматики
5. Контекстно-свободные грамматики
6. Нормальная форма Хомского КС-грамматик

Разнообразие формальных языков

машины Тьюринга

рекурсивно перечислимые языки

Широкий класс языков

Алгоритмическая неразрешимость задач анализа

??? **Широкий класс языков**

Быстрые алгоритм анализа

Быстрые алгоритмы анализа

Узкий класс языков

конечные автоматы

регулярные языки

Формальные грамматики

В 1957 г. американский лингвист Ноам Хомский (Noam Chomsky) опубликовал книгу «Syntactic Structures», в которой впервые предложил формальный подход к изучению структуры языков на основе строгого определения грамматик, порождающих языковые конструкции — словосочетания, предложения, фразы.

Эта работа радикально изменила взгляды лингвистов на способы описания и изучения естественных языков и оказалась чрезвычайно важной для создания и использования искусственных языков — программирования, описания данных, разметки текстов и пр.

Формальные грамматики

Идея Хомского.

Как описать множество грамматически правильных конструкций языка?

1. Нужно ввести грамматические понятия, обозначающие типы языковых конструкций.
2. Нужно определить грамматические правила, раскрывающие содержание одних грамматических понятий в терминах других грамматических понятий и слов описываемого языка.

Так возникли формальные грамматики, позволившие создавать математические методы и алгоритмы для решения задач синтаксического анализа и трансляции языков.

ФОРМАЛЬНЫЕ ГРАММАТИКИ

Пример.

Грамматические понятия: *sentence*, *noun-group*, *verb-group*, *noun*, *verb*, *adjective*, *adverb*.

Правила грамматики:

sentence ::= *noun-group verb-group* .
noun-group ::= *noun* | *adjective noun-group*
verb-group ::= *verb* | *adverb verb-group*
noun ::= **cats** | **bats** | **rats**
verb ::= **eat** | **like**
adjective ::= **black** | **white** | **big** | **small**
adverb ::= **easy** | **hardly** | **often** | **rare**

Грамматически правильные предложения:

black big cats often eat small white rats

black black bats often rare like big small cats

Формальные грамматики

Порождающая грамматика — это система $G = (\Sigma, \mathcal{N}, P, S)$, состоящая из

- конечного алфавита Σ терминальных букв (терминалов),
- конечного алфавита $\mathcal{N}$ нетерминальных букв (нетерминалов), $\Sigma \cap \mathcal{N} = \varnothing$,
- конечного множества P грамматических правил (продукций) вида $\alpha \to \beta$, где $\alpha \in (\Sigma \cup \mathcal{N})^* \mathcal{N} (\Sigma \cup \mathcal{N})^*, \beta \in (\Sigma \cup \mathcal{N})^*$,
- начального нетерминала S, $S \in \mathcal{N}$.

ФОРМАЛЬНЫЕ ГРАММАТИКИ

В приведенном примере описана грамматика, в которой

$\Sigma = \{$ a, b, c, . . . , x, y, z, . $\}$

$\mathcal{N} = \{$ *sentence*, *noun-group*, *verb-group*, *noun*, *verb*, *adjective*, *adverb* $\}$

P состоит из правил

$$sentence \to noun\text{-}group\ verb\text{-}group\ \ .$$
$$noun\text{-}group \to noun$$
$$noun\text{-}group \to adjective\ noun\text{-}group$$

. . .

$$adverb \to \mathbf{rare}$$

Начальный нетерминал S — это *sentence*

ФОРМАЛЬНЫЕ ГРАММАТИКИ

Пример (более абстрактный).

Грамматика $G = (\{a, b, c\}, \{S, B, C, H\}, P, S)$, где множество P состоит из правил

$$S \to aSBC$$
$$S \to aBC$$
$$CB \to HB$$
$$HB \to HC$$
$$HC \to BC$$
$$aB \to ab$$
$$bB \to bb$$
$$bC \to bc$$
$$cC \to cc$$

ФОРМАЛЬНЫЕ ГРАММАТИКИ

Грамматическое правило $r : \alpha \to \beta$ определяет отношение непосредственной выводимости слов $\xrightarrow{r}$: для любой пары слов $\gamma', \gamma'' \in (\Sigma \cup \mathcal{N})^*$

$$\gamma' \xrightarrow{r} \gamma'' \Leftrightarrow \gamma' = \theta\alpha\eta \;\wedge\; \gamma'' = \theta\beta\eta.$$

Для грамматики $G = (\Sigma, \mathcal{N}, P, S)$ отношение непосредственной выводимости $\xrightarrow{G}$ — это объединение отношений непосредственной выводимости для правил грамматики:

$$\xrightarrow{G} = \bigcup_{r \in P} \xrightarrow{r} .$$

Рефлексивно-транзитивное замыкание отношения $\xrightarrow{G}$ обозначим записью $\xrightarrow{G}_* .$

Формальные грамматики

Грамматическим выводом в грамматике G называется всякая конечная последовательность слов в алфавите $\Sigma \cup \mathcal{N}$, связанных отношением непосредственной выводимости:

$$\gamma_1 \xrightarrow{G} \gamma_2 \xrightarrow{G} \gamma_3 \xrightarrow{G} \cdots \xrightarrow{G} \gamma_{k-1} \xrightarrow{G} \gamma_k,$$

Слово γ'' выводимо в грамматике G из слова γ', если $\gamma' \xrightarrow{G}_* \gamma''$.

Язык, порождаемый грамматикой G — это множество $L(G) = \{w : w \in \Sigma^*, S \xrightarrow{G}_* w\}$ терминальных слов, выводимых из стартового нетерминала S.

Иерархия Хомского

Хомский предложил следующую классификацию формальных грамматик в зависимости от того, какие правила разрешается использовать для грамматического вывода. Эта классификация формальных грамматик и языков получила название

Иерархия Хомского

Иерархия Хомского

Грамматика $G = (\Sigma, \mathcal{N}, P, S)$ называется

- неограниченной (типа 0), если в ней разрешены любые грамматические правила;
- контекстно-зависимой (типа 1), если в ней разрешены только грамматические правила вида $\theta A \eta \to \theta \alpha \eta$, где $A \in \mathcal{N}$, $\alpha \in (\Sigma \cup \mathcal{N})^+$, $\theta, \eta \in (\Sigma \cup \mathcal{N})^*$;
- контекстно-свободной (типа 2), если в ней разрешены только грамматические правила вида $A \to \alpha$, где $A \in \mathcal{N}$, $\alpha \in (\Sigma \cup \mathcal{N})^*$;
- регулярные (типа 3), если в ней разрешены только грамматические правила вида $A \to wB$ или $A \to w$, где $A, B \in \mathcal{N}$, $w \in \Sigma^*$.

Иерархия Хомского

Пример. Неограниченная грамматика, моделирующая машины Тьюринга:

$\Sigma = \{0, 1, \$, \#\}$, $\mathcal{N} = \{S, Q_1, Q_2\}$.

$S \to \$Q_1 0\#$,

$0Q_1 0 \to Q_2 01$,

$1Q_1 0 \to Q_2 11$,

$Q_1 10 \to 0Q_1 0$,

$Q_1 11 \to 0Q_1 1$,

$Q_2 00 \to 0Q_1 0$,

$Q_2 01 \to 0Q_1 1$,

$Q_2 1 \to 0$,

$\$Q_1 \to \$0Q_1 \qquad \$Q_2 \to \$0Q_2$

$Q_1\# \to Q_1 0\# \qquad Q_2\# \to Q_2 0\#$

ИЕРАРХИЯ ХОМСКОГО

Пример. Контекстно-зависимая грамматика.

$\Sigma = \{a, b, c\}, \ \mathcal{N} = \{S, B, C, H\}$,

$$S \to aSBC$$
$$S \to aBC$$
$$CB \to HB$$
$$HB \to HC$$
$$HC \to BC$$
$$aB \to ab$$
$$bB \to bb$$
$$bC \to bc$$
$$cC \to cc$$

Иерархия Хомского

Пример. Контекстно-свободная грамматика.

$\Sigma = \{x, y, 0, 1, +, \times\}$, $\mathcal{N} = \{S, V, C, A\}$,

$$S \to V$$
$$S \to C$$
$$S \to (SAS)$$
$$V \to x$$
$$V \to y$$
$$C \to 0$$
$$C \to 1$$
$$A \to +$$
$$A \to \times$$

Иерархия Хомского

Пример. Регулярная грамматика, моделирующая конечный автомат.

$\Sigma = \{a, b\}$, $\mathcal{N} = \{S, Q_1, Q_2, Q_3\}$,

$$
\begin{aligned}
S &\to aQ_1 \\
S &\to aS \\
S &\to bQ_2 \\
Q_1 &\to aQ_2 \\
Q_1 &\to aS \\
Q_2 &\to bQ_1 \\
Q_2 &\to \varepsilon
\end{aligned}
$$

Иерархия Хомского

Утверждение 7.1.

1. Каждая регулярная грамматика является контекстно-свободной,
2. Каждая контекстно-свободная грамматика без правил вида $A \to \varepsilon$ является контекстно-зависимой,
3. Каждая контекстно-зависимая грамматика является неограниченной.

А теперь разберемся с грамматиками каждого типа по отдельности.

РЕГУЛЯРНЫЕ ГРАММАТИКИ

Праволинейные грамматики состоят из правил вида $A \to wB$ или $A \to w$, где $A, B \in \mathcal{N}$, $w \in \Sigma^*$

.

Поскольку в правой части каждого правила присутствует не более одного нетерминала, который может содержаться только в правом конце слова, такие грамматики также называются праволинейными .

Зададимся вопросом: какие языки порождаются регулярными (праволинейными) грамматиками?

Регулярные грамматики

Пусть задана грамматика $G = (\Sigma, \mathcal{N}, P, S)$ и нетерминал N, $n \in \mathcal{N}$. Языком нетерминала N, $N \in \mathcal{N}$ в грамматике G называется множество терминальных слов $L_G(N) = \{w : w \in \Sigma^*, N \xrightarrow{G}_* w\}$, выводимых из этого терминала.

Утверждение 7.2. Каждый язык, порождаемый регулярной грамматикой, является регулярным.

Доказательство. Проведем для случая, когда регулярная грамматика не содержит правил $N \to \varepsilon$, в правой части которых стоит пустое слово.

РЕГУЛЯРНЫЕ ГРАММАТИКИ

Пусть задана регулярная грамматика $G = (\Sigma, \mathcal{N}, P, S)$. Рассмотрим произвольный нетерминал $N,\ N \in \mathcal{N}$, и все грамматические правила $N \to w_i A_i,\ 1 \leq i \leq k$, и $N \to u_j,\ 1 \leq j \leq m$.

Легко видеть, что справедливо равенство

$$L_G(N) = \bigcup_{i=1}^{k} w_i L_G(A_i) \cup \{u_1, \ldots, u_m\}.$$

Таким образом семейство языков $L_G(N),\ N \in \mathcal{N}$, является решением системы линейных уравнений над языками

$$X_N = \sum_{j=1}^{k} w_i X_{A_i} + u_1 + \cdots + u_m.$$

Но, как следует из утверждения 3.3, эта система уравнений имеет единственное решение, и этим решением является набор регулярных языков. QED

РЕГУЛЯРНЫЕ ГРАММАТИКИ

Задача 1. Какое значение для предложенного доказательства имеет допущение об отсуствии в грамматике правил с пустой правой частью?

Проведите доказательство утверждения 7.2 для регулярных грамматик произвольного вида.

Утверждение 7.3. Каждый регулярный язык порождается регулярной грамматикой.

Доказательство. Поскольку класс регулярных языков совпадает с классом автоматных языков, достаточно показать, что каждый автоматный язык порождается регулярной грамматикой.

РЕГУЛЯРНЫЕ ГРАММАТИКИ

Пусть $L = L(\mathcal{A})$, где $\mathcal{A} = (\Sigma^*, Q, q_0, F, T)$ — детеминированный конечный автомат.

Рассмотрим регулярную грамматику $G = (\Sigma, Q, P, q_0)$, в которой в роли нетерминалов выступают состояния автомата $\mathcal{A}$, а множество грамматических правил для каждого перехода $(q', a, q'') \in T$ автомата $\mathcal{A}$ содержит правило $q' \to aq''$, и для каждого финального состояния q, $q \in F$, содержит правило $q \to \varepsilon$.

Далее индукцией по длине слова $w = a_1 a_2 \dots a_n$ нетрудно показать, что автомат $\mathcal{A}$ имеет успешное вычисление

$$q_0 \xrightarrow{a_1} q_1 \xrightarrow{a_2} \cdots \xrightarrow{a_n} q_n$$

тогда и только тогда в грамматике $G = (\Sigma, Q, P, q_0)$ существует вывод

$$q_0 \xrightarrow{G} a_1 q_1 \xrightarrow{G} \cdots \xrightarrow{G_n} a_1 a_2 \dots a_n q_n \xrightarrow{G_n} a_1 a_2 \dots a_n.$$

QED

Регулярные грамматики

Таким образом, мы доказали, что справедлива

Теорема 7.4. Класс регулярных языков совпадает с классом языков, порождаемых регулярными грамматиками.

Этим и объясняет такое название рассмотренных грамматик.

НЕОГРАНИЧЕННЫЕ ГРАММАТИКИ

Неограниченные грамматики по своим порождающим возможностям равносильны машинам Тьюринга.

Утверждение 7.5. Для любой неограниченной грамматики G существует недетерминированная двухленточная МТ $\mathcal{M}$, порождающая все слова языка $L(G)$ и только эти слова.

Доказательство. Нужная МТ $\mathcal{M}$ работает так: записывает на ленте начальный нетерминал S и моделирует вывод по правилам грамматики G, чередуя ленты для записи слов в выводе. Как только будет получено терминальное слово МТ $\mathcal{M}$ останавливается и записывает его как результат на выходной ленте. QED

Можно доказать и большее.

Задача 2. Докажите, что для любой пары алфавитов Σ, Γ существует такая МТ $\mathcal{U}$, что

$$L(\mathcal{U}) = \{w\#P : w \in \Sigma^*, G = (\Sigma, \Gamma, P, S) - \text{грамматика}, \ w \in L(G)\}.$$

НЕОГРАНИЧЕННЫЕ ГРАММАТИКИ

Утверждение 7.6. Для любого рекурсивно перечислимого языка L существует такая грамматика G, для которой верно $L = L(G)$.

Доказательство. Самостоятельно.

Подсказка: см задачу 4.6 и приведенный ранее пример неограниченной грамматики.

Теорема 7.7. Язык является рекурсивно перечислимым тогда и только тогда, когда он порождается некоторой формальной грамматикой.

КОНТЕКСТНО-СВОБОДНЫЕ ГРАММАТИКИ

Грамматика $G = (\Sigma, \mathcal{N}, P, S)$ называется контекстно-свободной, если в ней разрешены только грамматические правила вида $A \to \alpha$, где $A \in \mathcal{N}$, $\alpha \in (\Sigma \cup \mathcal{N})^*$;

Пример. $G = (\{a, b\}, \{S\}, P, S)$, где

$$P = \{S \ \to \ aSb \mid bSa \mid SS \mid \varepsilon\}.$$

Пример. $G = (\{a, b\}, \{S, A, B\}, P, S)$, где

$$\begin{aligned} P \ = \ \{S \ &\to \ aSB \mid bSA \mid aSBS \mid bSAS \mid \varepsilon, \\ A &\to a, \\ B &\to b\} \end{aligned}$$

Контекстно-свободные грамматики

Именно контекстно-свободные грамматики удобно использовать для описания синтаксиса многих искусственнных языков.

«оператор» $\to$ «простой оператор» $|$
begin «составной оператор» **end**
«составной оператор» $\to \varepsilon$ $|$ «оператор»; «составной оператор»
«простой оператор» $\to$ «левая часть» := «правая часть»
«левая часть» $\to$ «переменная»
«правая часть» $\to$ «терм»
«переменная» $\to$ «буква» «строка»
«буква» $\to$ a $|$ b $|$ c $|$ d $|$ e
«строка» $\to \varepsilon$ $|$ «буква» «строка» $|$ «цифра» «строка»
«цифра» $\to$ 0 $|$ 1 $|$ 2 $|$ 3 $|$ 4 $|$ 5 $|$ 6 $|$ 7 $|$ 8 $|$ 9

КОНТЕКСТНО-СВОБОДНЫЕ ГРАММАТИКИ

Вначале научимся упрощать КС-грамматики, избавляясь от тех конструкций, которые затрудняют анализ грамматик. Грамматики G и G' называются эквивалентными, если $L(G) = L(G')$.

Утверждение 7.8. Для любой КС-грамматики $G = (\Sigma, \mathcal{N}, P, S)$ существует эквивалентная КС-грамматика, у которой правые части правил не содержат начального нетерминала.

Доказтельство. Введем новый нетерминал S_0, объявим его начальным и добавим в грамматику правило $S_0 \to S$. QED

КОНТЕКСТНО-СВОБОДНЫЕ ГРАММАТИКИ

Грамматические правила вида $N' \to N''$, где $N', N'' \in \mathcal{N}$, называются переименованиями.

Утверждение 7.9. Для любой КС-грамматики $G = (\Sigma, \mathcal{N}, P, S)$ существует эквивалентная КС-грамматика, не содержащая правил-переименований.

Доказательство. Для каждого переименования $N' \to N''$ и правила $N'' \to \alpha$ добавим к множеству P новое правило $N' \to \alpha$ (если оно там отсутствовало). Ясно, что последовательное применение правил $N \to \theta N' \eta$ и $N' \to N''$ равносильно применению правила $N \to \theta N'' \eta$.

Как только обнаружится, что новых правил добавить невозможно, удалим все переименования. Поскольку в каждом успешном выводе переименование работает в паре с каким-то правилом, такое удаление нзменяяете порождаемого грамматикой языка.

QED

КОНТЕКСТНО-СВОБОДНЫЕ ГРАММАТИКИ

Грамматические правила вида $N \to \varepsilon$ называются ε-правилами.

Утверждение 7.10. Для любой КС-грамматики $G = (\Sigma, \mathcal{N}, P, S)$ существует эквивалентная КС-грамматика $G' = (\Sigma, \mathcal{N}, P', S)$, которая не содержит правил $N \to \varepsilon$ для всех нетерминалов $N, N \neq S$, отличных от начального нетерминала.

Доказательство. Проводится аналогично доказательству теоремы об устранении ε-правил для автоматов: если в множестве P есть пара правил $N \to \varepsilon$ и $N' \to uNv$, то добавим правило $N' \to uv$.

Очевидно, последовательное применение правил $N' \to uNv$ и $N \to \varepsilon$ равносильно применению правила $N' \to uv$. Поэтому как только возможности для добавления новых правил исчерпаны, можно удалить все ε-правила $N \to \varepsilon$, где $N \neq S$. QED

Контекстно-свободные грамматики

Пример

$$\begin{array}{lcl} S_0 & \to & \varepsilon \mid aSBb \\ S & \to & \varepsilon \mid aSBb \\ A & \to & \varepsilon \mid bbSA \mid aaSS \\ B & \to & \varepsilon \mid bbSA \mid aaSS \end{array}$$

$$\begin{array}{lcl} S_0 & \to & \varepsilon \mid aSBb \mid aBb \mid aSb \mid ab \\ S & \to & aSBb \mid aBb \mid aSb \mid ab \\ A & \to & bbSA \mid aaSS \mid bbA \mid bbS \mid bb \mid aaS \mid aa \\ B & \to & bbSA \mid aaSS \mid bbA \mid bbS \mid bb \mid aaS \mid aa \end{array}$$

Контекстно-свободные грамматики

Таким образом, справедлива

Теорема 7.11. Для любой КС-грамматики G существует эквивалентная КС-грамматика G', которая не содержит

- начального нетерминала в правых частях правил,
- правил-переименований,
- ε-правил для всех нетерминалов, отличных от начального.

КОНТЕКСТНО-СВОБОДНЫЕ ГРАММАТИКИ

Из теоремы 7/11 следует

Теорема 7.12. Алгоритмически разрешима проблема включения слова в КС-язык: $w \in L(G)$.

Доказательство. Как показано в утверждениях 7.8-7.10, любую КС-грамматику можно привести к виду, указанному в Теореме 7.11. Рассмотрим какой-нибудь успешный вывод в этой КС-грамматике

$$\alpha_0 = S \xrightarrow{G} \alpha_1 \xrightarrow{G} \alpha_2 \xrightarrow{G} \cdots \xrightarrow{G} \alpha_{n-1} \xrightarrow{G} \alpha_n = w.$$

Этот вывод либо проводится за один шаг $S \to \varepsilon$, либо является монотонно «возрастающим», т.е. для любого $i, 1 \leq n$ строка α_i либо имеет большую длину, чем строка α_{i-1}, либо имеет ту же длину, но содержит большее число терминалов, чем строка α_{i-1}. Почему?

Значит, для проверки включения $w \in L(G)$ достаточно рассмотреть все выводы длины не превосходящей $2|w|$. QED

КОНТЕКСТНО-СВОБОДНЫЕ ГРАММАТИКИ

Чтобы решить проблему пустоты $L(G) = \varnothing?$ для КС-грамматик, введем отношение зависимости нетерминалов: нетерминал N' зависит от нетерминала N'' в грамматике G (обозначение $N'' \preceq_G N'$), если существует такое грамматическое правило $N' \rightarrow \theta N'' \eta$, в правой части которого содержится нетерминал N''. Рефлексивно-транзитивное замыкание отношения зависимости нетерминалов будем обозначать записью $\preceq_G^*$.

Нетерминал N в грамматике G будем называть

- завершаемым, если $N \xrightarrow{G}_* w$ для некоторого терминального слова w;
- достижимым, если $S \xrightarrow{G}_* \theta N \eta$;
- полезным, если N присутствует в грамматическом выводе некоторого терминального слова из S.

КОНТЕКСТНО-СВОБОДНЫЕ ГРАММАТИКИ

Утверждение 7.13. Если нетерминал N не является полезным в грамматике G, то, удалив из нее все правила, содержащие N, получим грамматику, эквивалентную G.

Утверждение 7.14. Нетерминал является N достижимым в грамматике G тогда и только тогда, когда $N \preceq_G^* S$.

КОНТЕКСТНО-СВОБОДНЫЕ ГРАММАТИКИ

Утверждение 7.15. Множество завершаемых нетерминалов грамматики G — это наименьшее множество $\widehat{\mathcal{N}}$, удовлетворяющее следующим двум условиям:

1. если грамматика G содержит правило $N \to w$, где $w \in \Sigma^*$, то $N \in \widehat{\mathcal{N}}$,
2. если грамматика G содержит правило $N \to \alpha$ и все нетерминалы строки α содержатся в множестве $\widehat{\mathcal{N}}$, то $N \in \widehat{\mathcal{N}}$.

Доказательство. Самостоятельно.
(с применением индукции по длине грамматического вывода)

КОНТЕКСТНО-СВОБОДНЫЕ ГРАММАТИКИ

Пример

Грамматика G:

$$
\begin{array}{lcl}
S & \to & UX \mid VZ \\
T & \to & aav \mid bb \\
U & \to & aUa \mid bUb \\
V & \to & aTb \mid bTa \\
W & \to & YZY \mid aab \\
X & \to & \varepsilon \mid Xa \mid Xb \\
Y & \to & \varepsilon \mid YY \mid aU \\
Z & \to & W \mid b
\end{array}
$$

Достижимые нетерминалы: S, U, X, V, Z, T, W, Y .

Завершаемые нетерминалы: T, W, X, Y, Z, V, S .

Значит, U — бесполезный нетерминал, и его можно удалить.

КОНТЕКСТНО-СВОБОДНЫЕ ГРАММАТИКИ

Пример

Грамматика G':

$$
\begin{array}{lcl}
S & \to & VZ \\
T & \to & aa \mid bb \\
V & \to & aTb \mid bTa \\
W & \to & YZY \mid aab \\
X & \to & \varepsilon \mid Xa \mid Xb \\
Y & \to & \varepsilon \mid YY \\
Z & \to & W \mid b
\end{array}
$$

Достижимые нетерминалы: S, V, Z, T, W, Y .

Значит, X — бесполезный нетерминал, и его можно удалить.

КОНТЕКСТНО-СВОБОДНЫЕ ГРАММАТИКИ

Пример

Грамматика G'':

$$
\begin{array}{lcl}
S & \to & VZ \\
T & \to & aa \mid bb \\
V & \to & aTb \mid bTa \\
W & \to & YZY \mid aab \\
Y & \to & \varepsilon \mid YY \\
Z & \to & W \mid b
\end{array}
$$

Достижимые нетерминалы: S, V, Z, T, W, Y .

Завершаемые нетерминалы: T, W, Y, Z, V, S .

Все они полезные, и при этом $L(G'') = L(G)$.

КОНТЕКСТНО-СВОБОДНЫЕ ГРАММАТИКИ

Задача 7.4 Сформулируйте необходимое и достаточное условие полезности нетерминала в терминах свойств завершаемости и достижимости.

Задача 7.5 Опишите алгоритм вычисления всех полезных терминалов, обоснуйте его корректность и оцените его сложность.

КОНТЕКСТНО-СВОБОДНЫЕ ГРАММАТИКИ

Теорема 7.16. Алгоритмически разрешима проблема пустоты КС-языков: $L(G) = \varnothing$

Доказательство. Достаточно убедиться, что начальный нетерминал грамматики является завершаемым.

Нормальная форма Хомского

А нельзя ли унифицировать форму правил КС-грамматик?

Грамматика G в нормальной форме Хомского — это КС-грамматика, все правила которой имеют вид $S \to \varnothing$, $N \to a$ или $N \to N'N''$, где $N', N'' \in \mathcal{N} \setminus \{S\}$.

Пример. G: в нормальной форме Хомского.

$$
\begin{array}{lcl}
S & \to & \varepsilon \mid VZ \\
Z & \to & AA \mid BB \\
V & \to & a \mid ZV \\
A & \to & a \\
B & \to & b
\end{array}
$$

Нормальная форма Хомского

Теорема 7.17. Каждая КС-грамматика эквивалентная некоторой грамматике в нормальной форме Хомского

Доказательство. Опишем последовательность эквивалентных преобразований, приводящую произвольную КС-грамматику к нормальной форме Хомского.

1. Устраним начальный нетерминал S из правых частей правил (см. Утверждение 7.8);
2. Для каждого терминал a заменим его во всех правилах новым нетерминалом T_a и добавим новое правило $T_a \to a$;

Нормальная форма Хомского

3. Устраним ε-правила (см. Утверждение 7.10);
4. Устраним правила-переименования (см. Утверждение 7.9);
5. Вводя новые вспомогательные нетерминалы, приведем все нетерминальные строки в правых частях правил, к двучленному виду:

$$A \to BCDE \quad \Rightarrow \quad \begin{cases} A \to BA', \\ A' \to CA'', \\ A'' \to DE, \end{cases}$$

Каждое из преобразований — эквивалентное. После их применения образуется грамматика в нормальной форме Хомского. QED

КОНТЕКСТНО-СВОБОДНЫЕ ГРАММАТИКИ

Пример.

Пусть дана грамматика G:

$$
\begin{array}{lcl}
S & \to & VZ \\
T & \to & aa \mid bb \\
V & \to & aTb \mid bTa \\
W & \to & YZY \mid aab \\
Y & \to & \varepsilon \mid YY \\
Z & \to & W \mid b
\end{array}
$$

Очевидно, начального нетерминала нет в правых частях правил.

КОНТЕКСТНО-СВОБОДНЫЕ ГРАММАТИКИ

Пример.

Введем нетерминалы T_a и добавим новые правила $T_a \to a$

$$
\begin{array}{lcl}
S & \to & VZ \\
T & \to & T_a T_a \mid T_b T_b \\
V & \to & T_a T T_b \mid T_b T T_a \\
W & \to & YZY \mid T_a T_a T_b \\
Y & \to & \varepsilon \mid YY \\
Z & \to & W \mid T_b \\
T_a & \to & a \\
T_b & \to & b
\end{array}
$$

КОНТЕКСТНО-СВОБОДНЫЕ ГРАММАТИКИ

Пример.

Избавимся от ε-правил

$$
\begin{array}{lcl}
S & \to & VZ \\
T & \to & T_aT_a \mid T_bT_b \\
V & \to & T_aTT_b \mid T_bTT_a \\
W & \to & YZY \mid T_aT_aT_b \mid YZ \mid ZY \mid Z \\
Y & \to & YY \mid Y \\
Z & \to & W \mid T_b \\
T_a & \to & a \\
T_b & \to & b
\end{array}
$$

Нетерминал Y становится незавершаемым.

КОНТЕКСТНО-СВОБОДНЫЕ ГРАММАТИКИ

Пример.

Грамматика упрощается

$$
\begin{aligned}
S &\to VZ \\
T &\to T_a T_a \mid T_b T_b \\
V &\to T_a T T_b \mid T_b T T_a \\
W &\to T_a T_a T_b \mid Z \\
Z &\to W \mid T_b \\
T_a &\to a \\
T_b &\to b
\end{aligned}
$$

КОНТЕКСТНО-СВОБОДНЫЕ ГРАММАТИКИ

Пример.

Избавимся от правил-переименований

$$
\begin{array}{lcl}
S & \to & VZ \\
T & \to & T_a T_a \mid T_b T_b \\
V & \to & T_a T T_b \mid T_b T T_a \\
W & \to & T_a T_a T_b \mid b \\
Z & \to & T_a T_a T_b \mid b \\
T_a & \to & a \\
T_b & \to & b
\end{array}
$$

Нетерминал W становится недостижимым.

КОНТЕКСТНО-СВОБОДНЫЕ ГРАММАТИКИ

Пример.

Приведем правила к стандартному виду

$$
\begin{array}{lcl}
S & \to & VZ \\
T & \to & T_a T_a \mid T_b T_b \\
V & \to & T_a V' \mid T_b V'' \\
V' & \to & T T_b \\
V'' & \to & T T_a \\
Z & \to & T_a Z' \mid b \\
Z' & \to & T_a T_b \\
T_a & \to & a \\
T_b & \to & b
\end{array}
$$

Получаем грамматику в нормальной форме Хомского.

КОНТЕКСТНО-СВОБОДНЫЕ ГРАММАТИКИ

Задача 7.5. Предложите еще какие-нибудь правила эквивалентных преобразований, при помощи которых можно упрощаять КС-грамматики. Воспользуйтесь ими для еще более полного упрощения грамматики, рассмотренной в предыдущем примере.

Задача 7.6. Является ли алгоритмически разрешимой следующая проблема: выяснить, является ли контекстно-свободным язык, порождаемый заданной неограниченной грамматикой.

КОНТЕКСТНО-СВОБОДНЫЕ ГРАММАТИКИ

Задача 7.7. [Трудная] Является ли алгоритмически разрешимой следующая проблема: выяснить, является ли контекстно-свободным язык, порождаемый заданной контекстно-зависимой грамматикой.

КОНТЕКСТНО-СВОБОДНЫЕ ГРАММАТИКИ

Задача 7.8. Грамматика в нормальной форме Грейбах — это контекстно-свободная грамматика, в которой каждое правило имеет один из следующих четырех видов:

$$A \to \varepsilon, \qquad A \to a, \qquad A \to aB, \qquad A \to aBC.$$

Доказать, что любая КС-грамматика эквивалентна некоторой грамматике в нормальной форме Грейбах.

КОНЕЦ ЛЕКЦИИ 7