Домашняя работа по ТРЯП №10

Автор - Айвазов Денис из 671 группы

9 ноября 2017

1 КСГ порождающая $\Sigma^*\backslash\{a^nb^nc^n|n \geq 0\}$

КСГ это совокупность всех box-ов:
$\boxed{S \to S_1|S_2|S_3|S_{41}|S_{42}|S_{51}|S_{52}|S_{61}|S_{62}}$ - это начало значит, что мы по сути соединили 9 грамматик, каждая из которых принимает свою часть нужного нам языка. Значит нам надол будет доказать каждое из порождений поотдельности и показать, что это все нужные варианты.
Далее порождения: $\boxed{Q \to a|b|c|QQ|\varepsilon}$ - эквивалентно Σ^*, $A \to a|AA|\varepsilon$ - эквивалентно a^*, $\boxed{B \to b|BB|\varepsilon}$ - эквивалентно b^*, $\boxed{C \to c|CC|\varepsilon}$ - эквивалентно *. В случаях после 3 слово выглядит так: $\omega = a^nb^mc^k;\ n{+}m{+}k > 0$

1. случай $\Sigma^*b\Sigma^*a$ - когда в слове хоть одно b идет раньше a. Такой язык порождает грамматика: $\boxed{S_1 \to QbQaQ}$
2. случай $\Sigma^*c\Sigma^*a$. Похожим образом: $\boxed{S_2 \to QcQaQ}$
3. случай $\Sigma^*c\Sigma^*b$. Аналогично: $\boxed{S_3 \to QcQbQ}$
4. (a) $n > m \geq 0; \forall k;$ - букв a больше чем букв b, в нужном порядке. $\boxed{S_{41} \to A_4C;\ A_4 \to aA_4b|aA}$ т.е порождение какого-либо слова будет иметь вид: $S \Rightarrow S_{41} \Rightarrow A_4C \Rightarrow^* a^iA_4b^iC \Rightarrow a^iaAb^iC \Rightarrow^* a^{i+j}b^ic^l$; где $i, l > 0;\ j \geq 0$. Как раз то, что нужно. Букв a больше чем букв b на j штук.

   (b) $n < m \geq 0; \forall k;$ $\boxed{S_{42} \to B_4C;\ B_4 \to aB_4b|bB}$ т.е порождение какого-либо слова будет иметь вид: $S \Rightarrow S_{41} \Rightarrow B_4C \Rightarrow^* a^iB_4b^iC \Rightarrow a^ibBb^iC \Rightarrow^* a^ib^{i+j}c^l$; где $i, l > 0;\ j \geq 0$. Как раз то, что нужно. Букв b больше чем букв a на j штук.
5. (a) $n > k \geq 0; \forall k;$ - букв a больше чем букв c
   $\boxed{S_{51} \to aS_{51}c|BcC}$ т.е порождение какого-либо слова будет иметь вид: $S \Rightarrow S_{51} \Rightarrow aS_{51}c \Rightarrow^* a^iS_{51}c^i \Rightarrow a^iBcCc^i \Rightarrow^* a^ib^lc^{i+j}$ где $i, l > 0;\ j \geq 0$

   (b) $n < k \geq 0; \forall k;$ - букв c больше чем букв a
   $\boxed{S_{52} \to aS_{52}c|AaB}$ т.е порождение какого-либо слова будет иметь вид: $S \Rightarrow S_{52} \Rightarrow aS_{52}c \Rightarrow^* a^iS_{52}c^i \Rightarrow a^iAaBc^i \Rightarrow^* a^{i+j}b^lc^i$ где $i, l > 0;\ j \geq 0$
6. (a) $m > k \geq 0; \forall n;$ аналогично 4.1 $\boxed{S_{61} \to AB_6;\ B_6 \to bB_6c|Bb}$; Порождение: $S_{61} \Rightarrow AB_6 \Rightarrow^* Ab^iB_6c^i \Rightarrow Ab^iBbc^i \Rightarrow a^lb^{i+j}c^i$

(b) $m < k \geq 0; \forall n$; аналогично 4.2 $\boxed{S_{62} \to AC_6;\ C_6 \to bC_6c|cC}$
Порождение: $S_2 \Rightarrow A_6 \Rightarrow^* Ab^i{}_6c^i \Rightarrow Ab^ic^i \Rightarrow a^lb^ic^{i+j}$

Таким образом мы построили грамматику как объединение нескольких грамматик, которые разбирают все возможные варианты (кроме $a^nb^nc^n$, который нам как раз надо исключить).

2 Верно ли, что $L = \{a^nb^mb^nc^m | n, m \geq 0\} \subset CFL$

Да, он КС язык. Докажем построением и обоснованием КС-грамматики. Но для начала заметим, что $\forall n, m \Rightarrow \omega = a^nb^mb^nc^m \equiv a^nb^{m+n}c^m \equiv a^nb^nb^mc^m \equiv (a^nb^n)(b^mc^m) = \omega_1\omega_2$
Тогда сама грамматика: $S \to S_1S_2$; $S_1 \to aS_1b|\varepsilon$; $S_2 \to bS_2c|\varepsilon$; Видно, что любое порождение грамматики выглядит так: $S \Rightarrow S_1S_2 \Rightarrow^k a^kS_1b^kS_2 \Rightarrow a^kb^kS_2 \Rightarrow^j a^kb^kb^jS_2c^j \Rightarrow a^kb^kb^jc^j \equiv a^kb^jb^kc^j \quad \forall k, j \geq 0$ Т.е. мы получим нужный нам язык: $L = \{a^nb^mb^nc^m | n, m \geq 0\} \subset CFL$

3 $A, B \in CFL;\ R \in REG$ верно ли:

3.1 $A \backslash R \in CFL$

В множествах работает равенство: $A \backslash R \equiv A \cap \overline{R}$. А т.к. Регулярные языки замкнуты относительно операции дополнения, то $R_2 = \overline{R} \in REG$; $\Rightarrow$ наша задача свелась к доказательству того, что пересечение КС языка и регулярного языка - КС язык. Но это уже было доказано мной в 671Aivazov8 через конструкцию произведения. Для удобства копирую доказательство сюда:

Работает оно точно так же, как и конструкция произведения у ДКА. Будем считать, что регулярный язык задан ДКА $R = \{Q_R, \Sigma, q_0^R, \delta_R, F_R\}$, а КС задан МП-автоматом $M = \{Q_M, \Sigma, \Gamma, \delta_M, q_0^M, z_0, F_M\}$. Будем считать, что он принимает по принимающим состояниям (эквивалентность с автоматом, принимающим по пустому магазину в Хокрофт, Мотвани, Ульман на стр. 251). Тогда аналогично конструкции произведения с 3ого семинара можно построить МП автомат P, принимающий пересечение $L(M) \cap L(R)$ по правилам:
Состояния нового автомата это $Q_P = \{(Q_R, Q_M)\}$ - т.е. упорядоченные пары из декартова произведения множеств состояний ДКА и МП. Начальное состояние $q_0^P = (q_0^R, q_0^M)$. Алфавит Σ у всех один. Магазин тоже один. Функцию переходов δ_P определим так: $\delta((q_R, q_M), \sigma, Z) = (\delta_R(q, \sigma), \delta_M(Q, \sigma, Z)$. По сути мы ходим параллельно по двум автоматам, как и в конструкуции произведения. Принимающие состояния это $F_P = (F_R, F_Q)$ - если мы попали в состояние, эквивалентное принимающим у двух языков, значит слово было принято автоматом, т.е. принято его Регулярной часть и принято его Контекстно свободной частью в один момент времени.
Мы показали алгоритм построения МП-автомата, задающего пересечение REG и КС языков. $P = \{(Q_R, Q_M), \Sigma, \Gamma, \delta_P, (q_0^R, q_0^M), Z_0, (F_R, F_Q)\}$

Вердикт: разность КС языка и регулярного языка это КС язык, значит $A \backslash R \in CFL$.

3.2 $R \backslash A \in CFL$

Возьмем языки: $R = \{a,b,c\}^* \in REG$; $\quad L = \{a,b,c\}^* \backslash \{a^n b^n c^n | n \geq 0\} \in CFL$ (было доказано).
Тогда $R \backslash A = \{a,b,c\}^* \backslash \big(\{a,b,c\}^* \backslash \{a^n b^n c^n | n \geq 0\}\big) = \{a^n b^n c^n | n \geq 0\} \notin CFL$ (тоже было доказано на семинаре).

3.3 $A^R \in CFL$

Будем отражать (разворачивать в обратную сторону все правые части правил выводов грамматики A (т.к. язык A - КС, то существует грамматика, его порождающая).
Математической индукцией по длине вывода слова ω из грамматики A докажем что $\omega \in A \Leftrightarrow \omega^R \in A^R$.
База индукции: т.к. слово $\omega \in A$, то существует вывод из какого-то нетерминала S, т.е. где-то было правило $S \to \omega$, а значит и вывод $S \Rightarrow_A \omega$.
Предположение индукции: Пусть был вывод $S \Rightarrow_A^* \omega$ длины $\leq n$. Тогда предположим, что у нас есть вывод $S \Rightarrow_{A^R}^* \omega^R$
Шаг индукции: существует разбиение слова ω вида: $\omega = \omega_1 \omega_2 \omega_3 \dots \omega_k$ на k>1 слов, где хотя бы два слова не пустые. Тогда $\forall i$ выполнено: длинна вывода ω_i должна быть $< n$. А т.к. по предположению индукции для любого вывода меньшей длины мы можем разворачивать правила т.е. из $\forall i \;\; \exists$ вывод $S_i \Rightarrow_A^* \omega_i$ следует легальность вывода $S_i \Rightarrow_{A^R}^* \omega_i^R$. В итоге мы получаем, что $S \Rightarrow_{A^R}^* \omega^R$. Шаг доказан.
Т.е. по индукции доказано, что мы можем разворачивать в обратном порядке правые части правил, и $A^R \in CFL$

4 Доказать, что $L = \{\omega t \omega^R | \ |\omega| = |t|\} \notin CFL$

Докажем что язык не контекстно свободный от противного по Лемме о Накачке(ЛоН) для КСЯ.
Допустим, что $L \in CFL$. Тогда $\forall p \in \mathbb{N} \ \ \exists z \in L : z = \omega t \omega^R : |\omega| = |t| = |\omega^R| = p$; Тогда $\forall$ разбиения $z = uvwxy : \ |vwx| \leq p \ \& \ vx \neq \varepsilon$ разберем возможные случаи для подслова vwx:

1. $vwx = \omega[i, j]$, где $j - i + 1 \leq p$ - т.е. подслово ω с i-ой буквы по j-ую. Тогда по ЛоН $\forall i \geq 0$ слово $uv^iwx^iy \in L$. Определим $\omega' = \omega[1, i-1]u^iwx^i$. но для, например, i=0 получим $\omega' t \omega^R \notin L$ т.к. по условию $|\omega| = |t| = |\omega^R| > |\omega'|$ ведь $vx \neq \varepsilon$ (мы удалили минимум одну букву из начала). Получено противоречие с ЛоН, значит в данном случае язык не КС.

2. $vwx = t[i, j]$, где $j - i + 1 \leq p$
   Полностью аналогично, только минимум одну букву удалили из середины.

3. $vwx = \omega^R[i, j]$, где $j - i + 1 \leq p$. Точно так же доказывается.

4. $vwx = \omega[i, p] \cdot t[1, j]$, т.е. слово состоящее из конца ω и начала t причем суммарная длина их меньше р. $p - i + 1 + j + 1 \leq p$
   Похожим образом, только теперь по ЛоН $\forall i \geq 0$ слово $uv^iwx^iy \in L$. Определим $\omega' t' = \omega[1, i-1]u^iwx^it[j+1, p]$. Заметим, что теперь при $i \neq 1$ получаем что $|\omega' t'| \neq |\omega t|$ причем т.к. в данном случае мы убирали буквы из обоих слов, то выполнено что $|\omega'| \neq |\omega^R|$, а значит $\omega' t' \omega^R \notin L$, хотя по ЛоН $\omega' t' \omega^R \in L$ - противоречие.

5. $vwx = t[i, p] \cdot \omega^R[1, j]$ - аналогично

5 Являются ли CFL языки? А они $\subset REG$?

5.1 $L = \{xy : |x| < |y|, b \in x\}$

Построим КСГ. Слова x и y до последнего разделяют нетерминалы:

1. $Q \to a|b$ - символ означающий любую букву

2. $S \to aSQ|S_1$ - наполняем x буквами a, а y любыми. Пока $|x| = |y|$

3. $S_1 \to bS_2QQ$ - вот нам и встретилась буква b в слове x. И в этот же момент стало: $|x| = |y| - 1$

4. $S_2 \to QS_2Q|\varepsilon|Q$ - переход в QS_2Q говорит, что разность длин остается. А когда мы оказываемся в Q, то можно считать что с x-ом мы закончили, и все остальные буквы дописываем в y т.е. $|x| < |y|$

Регулярность языка следует из леммы о накачке для регулярных языков: мы просто "накачиваем"x так, что bшка уходит в часть слова, где y (иначе просто перестанет соблюдаться условие $|x| < |y|$)/ Т.е.: возьмем, например слово: $a^p b a^{p+m} \in L$ для $m \geq 2$ оно принадлежит языку, т.к. мы можем считать, что $x = a^p b$, а $y = a^{p+m}$ и $|x| = p + 1,\ |y| = p + m \geq p + 2 > p + 1 = |x|$.
Возможны случаи разбиения $\omega = uvw, |uv| \leq p, |u| > 0$:

1. $b \in u$. Тогда если накачивать слово, то при i=0 слово не подойдет языку, ведь теперь $|x| = |y|$.

2. $b \in v$. Тогда опять же при $i = 0$ у нас просто пропадает буква b и слово снова не подходит.

3. $b \in w$. Тогда начиная с какого-то момента накачивания $i \geq m$ получится, что (в зависимости от того, как посмотреть) либо b убежало из x, либо x стало длиннее y. В любом случае $\omega = a^{i+p} b a m + p \notin L$

Мы рассмотрели все разбиения. Доказано, что L принадлежит КСЯ и не принадлежит REG.

5.2 $L = \{xy : |x| < |y|, a \in y\}$

Регулярное выражение для этого языка выглядит так: $\Sigma^* a \Sigma^*$, где $\Sigma = \{a, b\}$. Это работает так: Мы можем называть x-ом, например, первый символ, а за y брать все остальное. И так увеличивать длину x пока не нарушится хотя бы одно из условий. Например разделения слова bbbbbbbbbbbbaababababaa: от $\underbrace{b}_{x}\underbrace{bbbbbbbbbbbaababababaa}_{y}$ и вплоть до $\underbrace{bbbbbbbbbb}_{x}\underbrace{bbaababababaa}_{y}$. А т.к. Регулярные языки вложены в кс языки, то этот язык кс. (если в это доказательство не верите, то) его можно задать КСГ: $S \to QaQ; Q \to a|b|\varepsilon|QQ \quad \equiv (a|b)^* a (a|b)^*$