ТРЯП, 9-ое Домашнее Задание

Сергей Пучинин, 873

25 ноября 2019 г.

Задача 1. Верно ли, что язык L является КС-языком? В случае положительного ответа построить КС-грамматику или МП-автомат для данного языка.

А. $L = \{a^n b^m b^n c^m \mid n, m \geqslant 0\}$, $\Sigma = \{a, b\}$;

Решение. Верно. Приведём соответствующую КС-грамматику:

$$S \vdash AC;$$
$$A \vdash aAb \mid \varepsilon,\ C \vdash bCc \mid \varepsilon.$$

Докажем соответствие данной КС-грамматике данному языку. Заметим, что $a^n b^m b^n c^m = a^n b^n b^m c^m$, то есть $L = RQ$, где

$$R = \{a^n b^n \mid n \geqslant 0\},\ Q = \{b^m c^m \mid m \geqslant 0\}.$$

Но заметим, что A порождает в точности R, а C порождает в точности Q. Следовательно, S задаёт L. Что и требовалось.

Б. $L = \{w : |w|_a \geqslant |w|_b \geqslant |w|_c\}$, $\Sigma = \{a, b, c\}$.

Решение. Не верно. Докажем это.

Будем доказывать от противного. Пусть $L \in \mathsf{CFL}$. Тогда для него выполнена Лемма о Накачке. Рассмотрим p из Леммы о Накачке. Возьмем слово $a^p b^p c^p$. Оно, очевидно, принадлежит L. Заметим, что для него в uv одновременно не могут входить и буквы a, и буквы c, так как $uv \leqslant p$, а расстояние между ближайшими буквами a и c равно p буквам.

Рассмотрим несколько случаев:

- В uv не входит буква c, но входит буква a. Тогда в слове xu^0yv^0z букв c — p, а букв a — не больше, чем $p-1$. Значит, данное слово не принадлежит L. Противоречие.

- В слове uv есть буква c и нет буквы a. Тогда в слове xu^2yv^2z букв a — p, а букв c — хотя бы $p+1$. Значит, данное слово не принадлежит L. Противоречие.

- Слово uv состоит только из букв b (при том, так как $uv \geqslant 1$, хотя бы одна буква b есть). Тогда в слове xu^2yv^2z букв a — p, а букв b — хотя бы $p+1$. Значит, данное слово не принадлежит L. Противоречие.

Получаем, что Лемма о Накачке не может выполняться для данного языка, следовательно, он не КС-язык.

Задача 2. Докажите, что язык $L = \{wtw^R \mid |w| = |t|\} \subseteq \{a, b\}^*$ не является КС-языком.

Доказательство. Для начала заметим, что так как $|w^R| = |w| = |t|$, то длина w линейно зависит от длины самого слова.

Еще заметим, что при возведении u и v в степени (в Лемме о Накачке), получаемые слова всегда начинаются на xu, а заканчиваются на vz.

Будем доказывать от противного. Пусть $L \in \mathsf{CFL}$. Тогда для него выполнена Лемма о Накачке. Возьмем p из Леммы о Накачке и рассмотрим следующее слово $a^p b^p a (ba)^{p-1} b b^p a^p$.[1] Оно, очевидно, принадлежит языку. Пусть u и v — подслова этого слова из Леммы о Накачке.

Рассмотрим несколько случаев[2]:

- $u \subset a^p$ (слева). Тогда v не затрагивает букв из a^p (справа). Тогда возведение u и v в степени будет увеличивать количество букв a с левого края, при этом не изменяя количество букв a с правого края.

- $u \subset b^p$ (слева). Тогда v не затрагивает букв из $abb^p a^p$ (справа). Тогда при возведении u и v в степени, количество букв b с левого края будет увеличивать, а с правого нет.

[1] Если степень p для дальнейших рассуждений будет слишком мала, то просто симметрично увеличим в этом слове все степени так, чтобы дальнейшие рассуждения о “не затрагивание каким-то подсловом букв другого подслова” работали.

[2] Симметричные случаи рассматриваются аналогично, так как небольшая асимметрия слова не влияет на рассуждения.

- u “пересекает” границу a^p и b^p (слева). Тогда v не затрагивает букв из $b^p a^p$ (справа). Тогда при возведении u и v в степени слово всегда будет начинаться на $a^p b^k a$, где $k < p$, так как $u \geqslant p$ и u содержит хотя бы одну a. А заканчиваться на $b^{k+1} a^p$.

- Иначе, при возведение u и v в степени не будут меняться начало и конец слова, а именно — $a^p b^p a$ всегда будет началом, а $bb^p a^p$ — концом.

Следовательно, в любой случае при возведении u и v в степень (и увеличении длины слова) мы выйдем за пределы языка (согласно замечанию в начале доказательства). Противоречие. Следовательно, $L \notin \mathsf{CFL}$. ■

Задача 3. Верно ли, что если язык L^* является КС-языком, то и язык L является КС-языком?

Решение. Не верно. Приведём пример.

Возьмем произвольный не КС-язык (например, язык квадратов) и объедим его с регулярным языком, состоящим из всех слов-символов алфавита, которых еще нет в текущем языке (язык регулярен, так как конечен). Так как КС-языки замкнуты относительно операции объединения и разности с регулярными языками, то полученный язык также является не КС-языком. Но его итерация, так как в нём присутствуют все слова-символы алфавита, совпадает с Σ^*, то есть является КС-языком.

Задача 4. Докажите, что КС-языки замкнуты относительно операции подстановки. То есть при подстановки КС-языков $L_1, L_2, \ldots, L_k$ в КС-язык M получается КС-язык $\sigma(M)$.

Доказательство. Рассмотрим КС-грамматики для языков $M, L_1, \ldots, L_k$. Переименуем их нетерминалы так, чтобы не было пересечений и чтобы аксиомы обозначались как $S, S_1, \ldots, S_k$, соответственно. Тогда заменим в КС-грамматике для M все терминалы на аксиомы, соответствующих им языков, и добавим в грамматику все правилы из грамматик для языков $L_1, \ldots, L_k$.

Утверждается, что данная грамматика порождает $\sigma(M)$. Докажем это.

В соответствии с грамматикой для M, нашими заменами терминалов на аксиомы и переименованием нетерминалов для устранения пересечений, мы из нашей грамматики для $\sigma(M)$ можем получить все выражения вида $S_{w_1} \ldots S_{w_{|w|}}$, где $w \in M$, и только их. Далее, из каждого нетерминала S_{w_i} мы можем получить любое слово $t \in L_{w_i}$, и только их, так как были добавлены все правила вывода слов из L_{w_i} из S_{w_i} и переименованы нетерминалы для устранения пересечений. То есть, наша грамматика порождает в точности

$$\bigcup_{w \in M} L_{w_1} \cdot L_{w_2} \cdot \ldots \cdot L_{w_{|w|}} = \sigma(M),$$

что и требовалось. ■

Задача 5. Докажите, что КС-языки префиксно замкнуты, т. е. для любого КС-языка L справедливо $\mathrm{Pref}(L) \in \mathsf{CFL}$.

Доказательство. Для начала докажем две вспомогательные Леммы.

Лемма 1. КС-языки замкнуты относительно прямого гомоморфизма.

Доказательство. Рассмотрим произвольный КС-язык L. Возьмём КС-грамматику для этого языка. Заменим в ней все терминальные символы на $h(x)$, где x — терминальный символ. Тогда новая грамматика будет порождать все слова вида

$$h(w_1)h(w_2)\ldots h(w_{|w|}),$$

где $w \in L$, и только их, так как мы не меняли сами деревья выводов, а лишь выражения на листьях. □

Лемма 2. КС-языки замкнуты относительно обратного гомоморфизма.

Доказательство. Рассмотрим произвольный КС-язык L. Возьмём МП-автомат M, порождающий этот язык. Построим автомат M' для языка $h^{-1}(L)$.

Пусть

$$M = \Big\{\Sigma, \Gamma, Q, q_0, Z_0, \delta, F\Big\},\ M' = \Big\{\Sigma, \Gamma, Q', q_0', Z_0, \delta', F'\Big\}.$$

Тогда

$$Q' = \Big\{[q, x] \mid q \in Q;\ x \in \mathrm{Suf}\big(h(\sigma)\big),\ \sigma \in \Sigma\Big\},\ q' = [q_0, \varepsilon].$$

Таким образом, первая компонента состояния является состоянием M, а вторая — состоянием буфера.

$$F' = \Big\{[q, \varepsilon] \mid q \in F\Big\}.$$

$$\delta'\big((q,\varepsilon),\sigma,z\big) = \Big\{\Big[\big(q,h(\sigma)\big),z\Big] \mid q\in Q;\ \sigma\in\Sigma;\ z\in\Gamma\Big\}, \tag{1}$$

$$(p,\gamma)\in\delta(q,c,z),\ c\in\Sigma\cup\varepsilon \to \big([p,x],\gamma\big)\in\delta'\big([q,cx],\varepsilon,z\big) \tag{2}$$

Согласно (1), когда буфер пуст, автомат M' может считать символ σ и поместить в буфер $h(\sigma)$. А согласно (2), M' может имитировать поведения автомата M, используя буфер для подачи автомату M символов.

Так как

$$(q_0,h(w),Z_0)\vdash^*_M (p,\varepsilon,\gamma),$$

то из всего выше сказанного получаем, что

$$\big([q_0,\varepsilon],w,Z_0\big)\vdash^*_{M'} \big\{[p,\varepsilon],\varepsilon,\gamma\big\},$$

то есть автомат порождает язык $h^{-1}(L)$. □

Теперь вернёмся к задаче. Пусть

$$\widetilde{\Sigma} = \{\widetilde{\sigma}\mid\sigma\in\Sigma\}.$$

Введём два отображения:

$$h:\ (\Sigma\cup\widetilde{\Sigma})^*\to\Sigma^*;\ \forall\sigma\in\Sigma\ h(\sigma)=\sigma,\ h(\widetilde{\sigma})=\sigma$$

и

$$g:\ (\Sigma\cup\widetilde{\Sigma})^*\to\Sigma^*;\ \forall\sigma\in\Sigma\ g(\sigma)=\sigma,\ g(\widetilde{\sigma})=\varepsilon.$$

Как легко видеть, данные отображения являются гомоморфизмами. Смысл первого — удаление волн над символами в слове, второго — удаление символов с волной из слова.

Поймём, что такое $h^{-1}(w)$. Это будет множество всех слов в алфавите $\Sigma\cup\widetilde{\Sigma}$, в которых символы идут в той же последовательности, что и в w, если забыть о волнах над ними.

Отсюда легко видеть[3], что

$$\mathrm{Pref}(L) = g\Big(h^{-1}(L)\cap\Sigma^*\widetilde{\Sigma}^*\Big).$$

А язык, стоящий справа, в силу замкнутости КС-языков относительно пересечения с регулярными языками и двух доказанных выше Лемм, является КС-языком. Следовательно, $\mathrm{Pref}(L)\in\mathsf{CFL}$. ■

[3] Для любого $u\in\mathrm{Pref}(L)$ существует слово $u\widetilde{v}$, где $v\in Suf(L)$, которое, очевидно, лежит в $h^{-1}(L)\cap\Sigma^*\widetilde{\Sigma}^*$. И среди слов из $h^{-1}(L)$ нас интересуют лишь те и только те слова, символы с волной в которых идут в конце.