Формальные языки

Домашнее задание 11

Фадеева Екатерина

2. S $\to$ aSbbbb | aaaSbb | c

Грамматика S $\to$ aSbbbb | T, T $\to$ aaaTbb | c эквивалентна исходной, т.к. в этой грамматике мы сначала делаем все правила вида S $\to$ aSbbbb из первой грамматики, а потом все правила вида S $\to$ aaaSbb из первой грамматики (а нам не важно в первой грамматике, в каком порядке выполнять все правила, кроме последнего правила S $\to$ c).

Любое слово, принадлежащее этой грамматике, имеет вид $a^n c b^m$, т.к. любое правило либо дописывает справа буквы b и слева буквы a, либо дописывает c и завершает разбор.

Любой вывод слова в этой грамматике имеет вид: выполняем x правил S $\to$ aSbbbb, затем правило S $\to$ T, затем y правил T $\to$ aaaTbb, затем правило T $\to$ c.

Теперь докажем, что приведенная грамматика однозначная: если у нас есть слово $a^n c b^m$ и его вывод с числами x и y, то известно, что выведется слово $a^{x+3y} c b^{4x+2y}$, а значит:

$$\begin{cases} x + 3y = n \\ 4x + 2y = m \end{cases}$$

Из этого можно однозначно вывести x и y:

$$\begin{cases} x = \frac{3m-2n}{10} \\ y = \frac{4n-m}{10} \end{cases}$$

Т.о. вывод слова $a^n c b^m$ — обязательно сначала выполняем $\frac{3m-2n}{10}$ правил S $\to$ aSbbbb, затем правило S $\to$ T, затем $\frac{4n-m}{10}$ правил T $\to$ aaaTbb, затем правило T $\to$ c, а значит вывод однозначен, чтд.

Т.о. ответ — S $\to$ aSbbbb | T, T $\to$ aaaTbb | c.

3. Это язык из слов таких, что:

1) Слово состоит из букв a и b
2) Общее количество букв a в слове в два раза больше общего количества букв b
3) На любом префиксе этого слова количество букв a не меньше, чем удвоенное количество букв b

$\Rightarrow$: любое слово из исходной грамматики соответствует правилам выше, это можно доказать по индукции по шагам вывода:

Если шаг вывода — F $\to \epsilon$, то правила выполняются

Если шаг вывода — F $\to$ aFaFbF, то по индукции для трех внутренних F правила выполняются, а значит верно, что в полученном в результате дальшейшего вывода слове общее число букв a в два раза больше, чем число букв b (мы добавили еще две буквы a и одну b), и неравенства на всех префиксах тоже выполняются (раз выполняются для любых префиксов внутренних F, и к ним добавляются либо одна a, либо две a, либо две a и одна b, а они ничего не испортят).

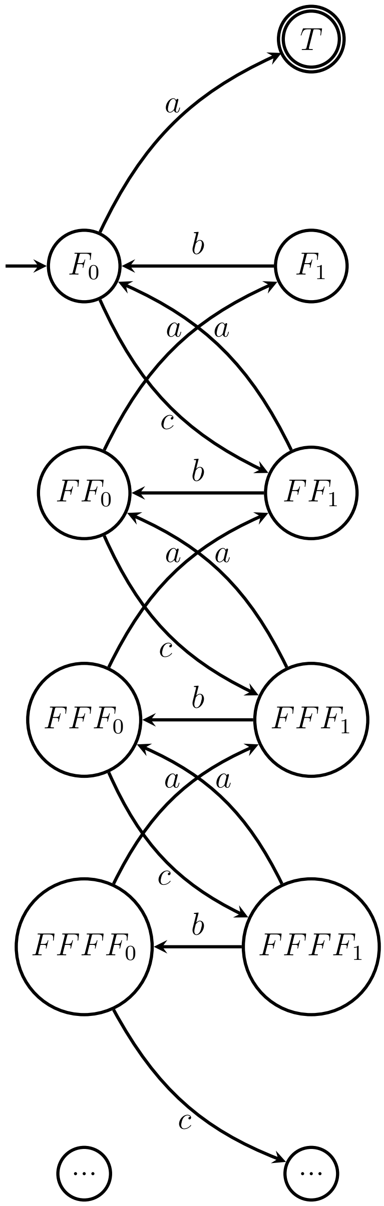
$\Leftarrow$: теперь покажем, что любое слово, удовлетворяющее правилам выше, выводится из F. Сделаем это по индукции от длины слова: пусть исходное слово — w. Если оно пустое, утверждение верно. Если же оно не пустое, выделим минимальный его непустой префикс $w[0..i]$, количество букв a на котором в два раза больше количества букв b (такой найдется, т.к. подходит вся строка). Этот префикс обязан начинаться с a и заканчиваться на b, т.е. w имеет вид $asbt$, где $asb = w[0..i]$. Для t выполняются три условия из ответа, тогда его можно по предположению индукции вывести из F. Осталось показать, что s можно вывести из FaF.

Раз префиксный балланс $asbt$ не меньше нуля, а общий балланс равен нулю, то (если a прибавляет 1 к общему баллансу, а b убавляет 2) общий баланс s равен 1, и раз asb — минимальная строка, балланс которой - 0, префиксные баллансы этой строки не меньше 1, а значит префиксные баллансы s не больше 0. Т.о. префиксные баллансы s не отрицательные, а общий балланс — 1. Тогда если ни один префиксный балланс s не равен нулю, то можно вывести s из FaF, если первое F привести к ϵ (s обязательно начинается с a, остаток — по индукции). Иначе если у s существует хотя бы один префиксный балланс 0, возьмем максимальный такой префикс, и выведем его из F (можно по индукции), затем обязательно должно следовать a, и остаток s можно опять вывести из F по индукции. Т.о. всегда можно вывести $FaF \to s$, а значит можно вывести $w = asbt$ из $aFaFbF$

4. Первая грамматика: F $\to$ a | bF | cFF

Вторую грамматику можно переписать: K $\to$ aaK | abK | caK | cbK | a | c

Можно нарисовать бесконечный автомат, который принимает пересечение этих двух языков (пояснение внизу):



Этот автомат — пересечение бесконечного автомата для первого языка и конечного автомата для второго:

1) первый столбец вершин — если мы находимся в K из второй грамматики, т.е. уже выписали четное число букв и можем ходить только по a или c;
2) второй столбец вершин — если мы находимся в M из второй грамматики, т.е. уже выписали нечетное число букв и можем ходить только по a или b.
3) строка грамматики F^n обозначает состояние, в котором мы уже выписали какой-то префикс грамматики, и осталось вывести какой-то остаток из нетерминала F n раз (в том числе вершина T — когда нам осталось вывести что-нибудь из нуля нетерминалов F, т.е. ничего; это вершиниа терминальная)

Например, вывод $F \to cFF \to cbFF \to cbcFFF \to cbcbFFF \to cbcbaFF \to cbcbaaF \to cbcbaaa$ соответствует пути в автомате $F_0 \to FF_1 \to FF_0 \to FFF_1 \to FFF_0 \to FF1 \to F_0 \to T$ (получается, состояние — количество F в конце + четность количества уже выведенных терминалов в начале)

По этому бесконечному автомату мы хотим построить КС грамматику. Для этого обозначим X - все пути из какой-то вершины слева в себя, при этом не переходя в строки выше исходной (т.е. в вершины с меньшим количеством F, чем в начале), Y - все пути из какой-то вершины справа в себя (тоже не заходящие выше). Вычислим X и Y рекурсивно:

Из автомата можно вывести, перебирая все пути:

$$\begin{cases} \text{X} \to \epsilon \mid \text{cYaX} \mid \text{cYbXaYbX} \\ \text{Y} \to \epsilon \mid \text{bXcYbXaY} \end{cases}$$

Терминальное состояние достижимо только из Xa.

Т.о. мой ответ:

$$\begin{cases} \text{S} \to \text{Xa} \\ \text{X} \to \epsilon \mid \text{cYaX} \mid \text{cYbXaYbX} \\ \text{Y} \to \epsilon \mid \text{bXcYbXaY} \end{cases}$$