Задачи на лемму о разрастании для регулярных языков

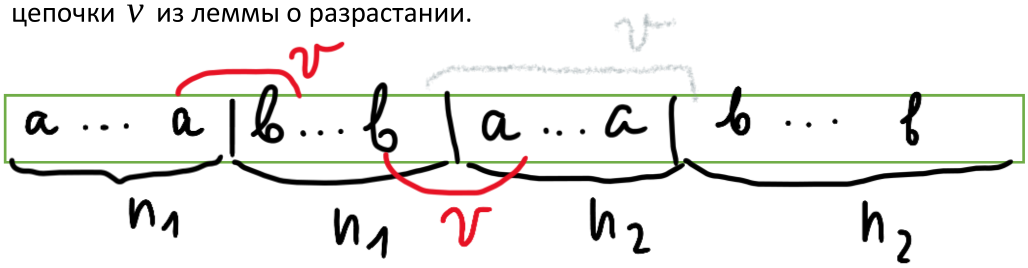
Помимо примеров, разобранных на лекции, рассмотрим некоторые более трудные задачи.

1) Язык $L = L_1^2 L_1^*$, где $L_1 = \{a^n b^n : n > 0\}$.

   Чтобы доказать нерегулярность такого языка, рассмотрим пересечение
   $L_2 = L \cap a^*b^*a^*b^* = \{a^{n_1}b^{n_1}a^{n_2}b^{n_2} : n_1, n_2 > 0\}$.

   Заметим, что в силу того, что пустая цепочка не принадлежит языку L_1, каждая цепочка языка L начинается префиксом вида $a^{n_1}b^{n_1}a^{n_2}b^{n_2}; n_1, n_2 > 0$, поэтому пересечение $L \cap a^*b^* = \varnothing$.

   Рассмотрим тогда достаточно длинную цепочку записанного выше пересечения L_2. На рисунке ниже показаны различные возможные способы расположения накачиваемой цепочки v из леммы о разрастании.



   Исключая очевидно невозможное расположение накачиваемой цепочки внутри какой-либо «зоны» (символов a или b), проанализируем накачку на стыках зон.

   Если $v = a^k b^l, 0 < k, l < n_i (i \in \{1,2\})$, получим, что $v^2 = a^k b^l a^k b^l$, то есть при накачке возникнут новые вхождения цепочки ba, что противоречит структуре цепочек языка L_2, в которых цепочка ba входит ровно один раз.

   Если же $v = b^l a^k, 0 < k, l < n_i (i \in \{1,2\})$, то $v^2 = b^l a^k b^l a^k$, и накачка приведет к появлению лишнего вхождения цепочки ab, которая в каждую цепочку языка L_2 входит ровно два раза.

   Расположение же накачиваемой «в обхват» зоны (серый цвет на рисунке) невозможно ввиду ограничения на длину накачиваемой цепочки: $|v| \le k_L$. Константа k_L, фиксируемая где-то на числовой прямой в силу предположения о регулярности анализируемого языка, может быть сколь угодно превзойдена выбором параметров n_1 и n_2 так, чтобы $n_1, n_2 > k_L$. Эту константу мы, разумеется, знать не можем, но само предположение о регулярности гарантирует существование этой константы как *фиксированной* точки на числовой прямой, которую можно превзойти, задав длины «зон» с учетом того, что длины цепочек анализируемого языка не ограничены сверху.

   Итак, язык L_2 нерегулярен, и вместе с ним нерегулярен и исходный язык L.

2) Язык $L = L_1^3 L_1^*$, где $L_1 = \{a^m b^n : m, n > 0, m \ne n\}$.

   Поскольку каждая цепочка этого языка начинается префиксом
   $a^{m_1}b^{n_1}a^{m_2}b^{n_2}a^{m_3}b^{n_3}; m_i \ne n_i, i \in \{1,2,3\}$,
   пересечем его с регулярным языком $(a^*b^*)^3$. В пересечении получим:

$L_2 = L \cap (a*b*)^3 = \{a^{m_1}b^{n_1}a^{m_2}b^{n_2}a^{m_3}b^{n_3}; m_i \neq n_i, m_i, n_i > 0; i \in \{1,2,3\}\}$,
а чтобы доказать нерегулярность полученного пересечения, рассмотрим язык
$L_3 = \overline{L}_2 \cap (a^+b^+)^3 = \{a^{n_1}b^{n_1}a^{n_2}b^{n_2}a^{n_3}b^{n_3}; n_i > 0, i \in \{1,2,3\}\}$.
Заметим, что выбор позитивной итерации в записанном выше выражении принципиален, так как в дополнение языка L_2 попадут цепочки вида
$a^{m_1}b^{n_1}a^{m_2}b^{n_2}a^{m_3}b^{n_3} (m_i, n_i \geq 0, i \in \{1,2,3\})$, то есть возможны пропуски некоторых зон, и соотношения между числами вхождений символов будет произвольным, так как неравенство $m_i \neq n_i$ выполняется только для двух соседних зон.
Нерегулярность языка L_3 доказывается совершенно аналогично доказательству нерегулярности языка L_2 в предыдущей задаче.

3) Пример кс-языка, не являющегося регулярным, но удовлетворяющего условию леммы о разрастании для регулярных языков:

$$L = \{a^n b^m a^n : m, n \geq 0\}.$$

Чтобы доказать, что язык удовлетворяет условию леммы о разрастании, необходимо указать выбор константы k_L, а затем проверить возможность накачки.

Здесь можно положить:

$$u = a^n b^k, v = b, w = b^l a^n, k + l + 1 = m,$$

если цепочка содержит хотя бы одну букву b; для цепочки a^{2n} можно считать $v = aa$. Это значит, что можно принять $k_L = 2$. Т.е., если $|x|=2$, то $x = aa$, а самая короткая цепочка, длины, не меньшей 2, с буквой b имеет вид aba.

Нерегулярность данного языка доказывается рассмотрением пересечения

$$L \cap a*ba* = \{a^n b a^n : n \geq 0\},$$

уже не удовлетворяющего условию леммы о разрастании.

4) В продолжение предыдущего примера рассмотрим язык:
$L = \{a^n b^m a^p : m, n \geq 0, p > n\}$.
Понятно, что сразу применить лемму о разрастании тут нельзя, так как невозможно будет отвергнуть возможность расположения накачиваемой цепочки в зоне символов b.
Образуем пересечение
$L_1 = L \cap a*ba* = \{a^n b a^p : n \geq 0, p > n\}$.
Все случаи расположения накачиваемой цепочки легко отвергаются, кроме одного – в правой зоне символов *a*. Но тут используем «прием выбрасывания» накачиваемой цепочки (ведь по лемме о разрастании ее можно не только повторять сколько угодно раз, но можно и выбросить). Тогда рассмотрим цепочку

$a^n b a^{n+1}$. Если положить $v = a^k, 0 < k < n+1$ и поместить ее в правую зону символов *a*, то после ее выбрасывания получим цепочку $a^n b a^{n+1-k}$. Так как $n+1-k \le n$, то такая цепочка уже не будет принадлежать языку L_1, который, следовательно, нерегулярен, а вместе с ним нерегулярен и исходный язык L. Можно показать, что и этот (исходный) язык удовлетворяет лемме о разрастании. Действительно, если $m > 0$, то можно положить $v = b$. Если же $m = 0$, то есть цепочка имеет вид $a^n a^{n+k} = a^{2n+k}, k > 0$, то можно положить $v = a^2$. При накачке получим цепочки вида $a^{2(n+m)+k}, m, k \ge 1$, которые принадлежат языку L. Значит, можно принять $k_L = 3$ (очевидно, самая короткая цепочка при $m > 0$ есть ba, а $a \in L$, но $\lambda \notin L$, и цепочку a^2 в качестве исходной при $v = a^2$ взять нельзя).

Задачи для самостоятельного решения

Доказать нерегулярность следующих языков:

1) $L = \{w^n : w \in V^*, |V| > 1, n > 1\}$ (n - фиксированное число).
2) $L = \{a^m b^n c^p : m, n, p \ge 0, m = p^2 + 1\}$.
3) Язык L определяется уравнением: $x = ab + axb + x^2$.
4) $L = \{x c^n y : x, y \in \{a, b\}, c \notin \{a, b\}, |x| = |y|, n > 0\}$.