Белоусов А.И.

УДК 519.76+372.851

О методике изложения некоторых разделов теории формальных языков: леммы о разрастании

al_belous@bk.ru

Введение

Теория формальных языков (ТФЯ) является в математической подготовке специалиста по программным технологиям одной из важнейших дисциплин. Дисциплина эта достаточно сложная, требующая тщательной методической проработки. К сожалению, в отечественной учебной литературе имеет место заметный дефицит пособий по ТФЯ. В [1] мы пытались восполнить этот дефицит, изложив достаточно подробно теорию регулярных и контекстно-свободных языков. Материал в соответствующих разделах основан на лекциях, долгое время читавшихся автором статьи студентам программистских специальностей.

В предлагаемой статье рассматриваются так называемые *леммы о разрастании* (или, *леммы о накачке*; в англоязычной литературе распространен термин the pumping lemmas). Название «леммы» исторически сложилось, но на самом деле это фундаментальные теоремы, дающие важные и, что существенно, конструктивно проверяемые необходимые условия принадлежности языка к тому или иному классу (регулярных, линейных, контекстно-свободных) языков. Тем самым именно с помощью этих теорем может быть выстроена иерархия классов языков, четкое представление о которой должен иметь любой квалифицированный программист.

Главной особенностью данной статьи является изложение весьма сложного доказательства леммы Огдена о контекстно-свободных языках, частным случаем которой является лемма о разрастании для этого класса языков. Насколько известно, эта лемма не получила отражения в отечественной учебной литературе, тогда как она позволяет анализировать такие языки, которые нельзя проанализировать с помощью леммы о разрастании.

В статье разбираются разные примеры применения леммы Огдена, лемм о разрастании для регулярных и КС-языков, такие, которые обычно не рассматриваются в известных нам руководствах (неравенства чисел вхождений символов, анализ языков, удовлетворяющих леммам, но не принадлежащих соответствующему классу языков). При решении задач полезно иногда прибегать к квазигеометрической иллюстрации, что облегчает понимание сути дела. Эта методика использована и в [1] . В примере на лемму Огдена вводится метод, который может быть назван «методом факториальной накачки» (см. также [2]).

Доказательство леммы Огдена основано на книгах [2] и [3] с устранением некоторых неточностей и восполнением пробелов.

Также дается подробное доказательство леммы о разрастании для линейных языков, обычно помещаемое в число задач для самостоятельного решения. Рассматриваются примеры применения этой леммы.

В последнем разделе мы даем пример доказательства, в котором лемма о разрастании используется для доказательства того, что язык принадлежит определенному классу языков, тогда как обычно эти леммы используются для негативных утверждений о непринадлежности языка тому или иному классу. Набросок такого доказательства можно найти в книге [4].

Излагаемый в статье материал ориентирован преимущественно на студентов-магистров программистских специальностей, а также на прикладных математиков, второго образования в том числе, изучающих дискретную математику и связанные с нею дисциплины (математическую логику, теорию алгоритмов, теорию графов).

1. Леммы о разрастании для регулярных и контекстно-свободных языков

В этом разделе мы приводим без доказательства леммы о разрастании для регулярных и контекстно-свободных (КС-) языков вместе с примерами их применения. Доказательства не приводятся, так как они во всех подробностях изложены в [1].

Теорема 1 (*Лемма о разрастании для регулярных языков*). Для любого регулярного языка L определена константа k_L (зависящая от L) такая, что всякая цепочка x языка L, длина которой не меньше k_L, представима в виде соединения трех цепочек: $x = uvw$, где цепочка v не пуста, ее длина не превосходит k_L, и для любого неотрицательного целого n цепочка uv^nw принадлежит языку L.

Теорема 2 *(Лемма о разрастании для КС-языков)*. Для любого КС-языка L определена константа k_L (зависящая от L) такая, что всякая цепочка z языка L, длина которой больше k_L, представима в виде соединения пяти цепочек: $z = uxwyv$, где цепочка w не пуста, цепочки x и y одновременно не пусты, длина цепочки xwy не превосходит k_L, и для любого неотрицательного целого n цепочка ux^nwy^nv принадлежит языку L.

Обсуждая эти леммы, полезно разобрать примеры анализа языков, не являющихся регулярными или контекстно-свободными. Здесь мы рассмотрим такие примеры, которым обычно не уделяется должного внимания, а именно, примеры языков, содержащих слова, в которых числа вхождений букв подчиняются определенным неравенствам.

Примеры. 1. Рассмотрим язык $L_1 = \{a^nb^m \colon n > m\}$ в алфавите $\{a, b\}$. Докажем, что он не регулярен.

Анализируя примеры подобного рода, полезно прибегнуть к некоторым наглядным иллюстрациям. Такая методика, принятая и в [1], как показывает опыт, способствует лучшему усвоению материала.

В данном случае можно наглядно представить произвольную (достаточно длинную) цепочку языка и возможные способы расположения в ней «накачиваемой» цепочки v, например, следующим образом (рис. 1):

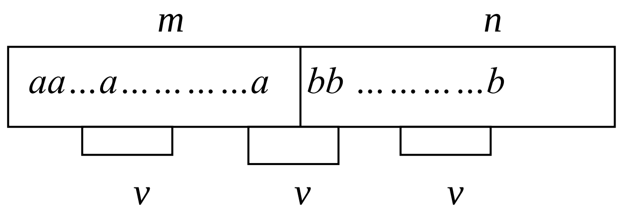

Рис. 1

Или так (рис. 2):

$$\underbrace{\overbrace{aa}^{v}\ldots aaa\ldots\overbrace{a\ldots ab\ldots b}^{v}}_{m}\underbrace{\overbrace{bb}^{v}\ldots bb\ldots b}_{n}$$

Рис. 2

Тогда можно в анализируемой цепочке можно говорить о «зонах»: зона символа a длины m, зона символа b длины n. И накачиваемая цепочка может располагаться целиком в зоне символов a, или целиком в зоне символов b, или, как можно неформально сказать, «на стыке» зон. Формально это выражается так: 1) $v = a^s$, где $0 < s < m$; 2) $v = b^s$, где

$0 < s < n$; 3) $v = a^s b^t$, где $0 < s < m$, $0 < t < n$.

Далее важно показать студентам, что решая задачу, нужно отвергнуть все возможные способы расположения накачиваемой цепочки. Если хотя бы один способ не будет доказательно отвергнут, задача не решена, и вопрос о регулярности или нерегулярности рассматриваемого языка остается открытым.

В данном примере все случаи расположения накачиваемой цепочки v легко отвергаются, кроме ее расположения в зоне символов a. Тогда ее можно повторять сколько угодно раз, не нарушая структуру слов языка. Но важно учесть, что в лемме о разрастании для регулярных языков допускается и выбрасывание накачиваемой цепочки. Рассмотрим тогда слово языка L_1 вида $a^{m+1}b^m$. Тогда, если $v = a^s$, то, поскольку цепочка v не пуста, $s > 1$, и, выбрасывая ее из исходного слова, получим слово

$a^{m+1\text{-}s}b^m$, не принадлежащее языку L_1. Следовательно, этот язык не регулярен.

2. Язык

$$L_2 = \{a^n b^m : n \neq m\}.$$

При анализе этого языка следует использовать алгебраические свойства множества регулярных языков, а именно, замкнутость его относительно операций пересечения и дополнения. Но если «прием пересечения» хорошо методически отработан [1], то реже используют другие свойства, в частности, замкнутость относительно дополнения.

Рассмотрим тогда язык

$$L_3 = \overline{L_2} \cap a^*b^*.$$

В предположении регулярности языка L_2 этот язык должен быть регулярным. Но нетрудно видеть, что пересекая дополнение языка L_2 с регулярным языком a^*b^*, мы оставим только слова вида $a^n b^n$, но язык, состоящих из всех слов такого вида, как известно, не регулярен.

Рассмотрим еще более трудную задачу на лемму о разрастании для регулярных языков.

3. Определим язык

$$L_4 = L_{21}^2 L_{21}^*,$$

где язык L_{21} определен точно так же, как и язык L_2, но только числа n и m должны быть положительными. Пересечем этот язык с регулярным языком $a^+b^+a^+b^+$ (заметим, что пересечение L_4 с a^+b^+ пусто в силу того, что числа n и m оба не равны нулю). В пересечении получим язык

$$L_{21}^2 = \{a^{m_1}b^{n_1}a^{m_2}b^{n_2} : m_i \neq n_i, m_i, n_i > 0\},$$

нерегулярность которого доказывается, как в предыдущей задаче рассмотрением пересечения

$$\overline{L_{21}^2} \cap a^+b^+a^+b^+ = \{a^m b^m a^n b^n : m, n > 0\}.$$

Нелишне заметить, что если в рассмотренном только что примере допустить нулевые числа n и m, то можно доказать, используя возможность пропуска одного из символов в каждой паре, что он регулярен и представляется выражением

$$(a^+b^* + a^*b^+)^* = L_2^2 L_2^*.$$

Примеры анализа языков с помощью леммы о разрастании для КС-языков, в которых также фигурируют неравенства, рассмотрены в [1] (пример 8.13).

Но вот возьмем язык

$$L_5 = \{a^n b^m c^p : n \neq m, n \neq p, m \neq p\},$$

в котором числа вхождений всех трех букв попарно различны. Анализ этого языка по лемме о разрастании для КС-языков был бы весьма затруднителен, и требуется более мощный аппарат, которым служит доказываемая ниже лемма Огдена. Этот язык будет проанализирован ниже.

Этот же раздел заключим весьма важным примером, в котором строятся языки, не являющиеся регулярными или контекстно-свободными, но при этом удовлетворяют указанным леммам. Важно обратить внимание студентов на это обстоятельство, подчеркнув еще раз, что леммы о разрастании дают лишь необходимые условия принадлежности языка соответствующему классу языков.

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

$$L_6 = \{a^n b^m a^n : m, n \geq 0\}.$$

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

$$L_6 \cap a^* b a^* = \{a^n b a^n : n \geq 0\},$$

уже не удовлетворяющего условию леммы о разрастании.

Рассмотрим теперь язык

$$L_7 = \{a^n b^m a^n b^p a^{n+1} : m, n, p \geq 0\}.$$

Докажем, что этот язык, не будучи контекстно-свободным, удовлетворяет лемме о разрастании для КС-языков. Для цепочки этого языка, у которой хотя бы одно из чисел m или p отлично от нуля, можно положить $x = b, w = a, y = \lambda$; если же $m = p = 0$, то для любой цепочки $a^{3n+1}, n > 0$, можно принять $x = aaa, v = a, y = \lambda$. Тем самым условия леммы о разрастании для КС-языков выполняются для любой цепочки, длина которой больше трех (т.е. константа k_L может быть приравнена 3). То, что язык L_2 не является контекстно-свободным, следует из невыполнения условий леммы о разрастании для пересечения L_7 с регулярным языком $a^*ba^*ba^+$.

Замечание. Напомним, что лемма о разрастании доказывается в предположении, что порождающая грамматика для КС-языка задана в приведенной форме, причем можно полагать, что аксиомы нет в правых частях правил вывода. Это значит, что цепочка w не пуста.

2. Лемма Огдена о КС-языках

Теорема 3 (*Лемма Огдена о КС-языках*). Для всякого КС-языка L существует константа k_L такая, что всякая цепочка z языка L, содержащая не менее k_L *выделенных позиций* (т.е. отмеченных вхождений символов терминального алфавита) представима в виде соединения пяти цепочек: $z = uxwyv$, где (1) цепочки $u,\ x$ каждая или $y,\ v$ каждая имеют выделенные позиции, (2) цепочка w содержит выделенную позицию, (3) цепочка xwyсодержит не более k_L выделенных позиций и (4) для всякого неотрицательного целого n цепочка $z_n = ux^nwy^nv$ принадлежит языку L.

Доказательство. Считаем, что кс-грамматика $G = (V,\ N,\ S,\ P)$, порождающая язык L, задана в приведенной форме, и правые части правил вывода не содержат вхождений аксиомы (т.е. правило $S \to \lambda$, если оно существует, используется исключительно для порождения пустой цепочки). Положим

$$m = |N|,\ l = \max_{A \to \alpha \in P} |\alpha|,$$

а константу определим так:

$$k_L = l^{2m+3}.$$

При доказательстве нам потребуется понятие максимального поддерева данного дерева. Поддерево с заданной корневой вершиной $\mathbf{T}$ называется максимальным, если в нем остаются все листья, достижимые из $\mathbf{T}$ в исходном дереве.

Рассмотрим теперь цепочку z языка L, содержащую не менее k_L выделенных позиций. Это значит, что крона дерева вывода (т.е., множество листьев дерева, упорядоченное слева направо) цепочки z содержит не менее k_L отмеченных (выделенных) листьев. В грамматике G максимальная степень ветвления любого дерева вывода составляет l (такое дерево называется, как известно, l *-деревом*), откуда следует, что дерево вывода высотой h содержит не более l^h листьев, или, что равносильно, дерево вывода, крона которого содержит не менее k_L листьев, имеет высоту, не меньшую $\log_l k_L$, т.е. при указанном определении константы k_L, высота дерева вывода цепочки z будет не меньше $2m+3$. Это значит, что в таком дереве существует путь из корня в лист, длина которого не меньше $2m+3$, а число вершин составит по крайней мере $(2m+3)+1$.

Обозначим дерево вывода цепочки z буквой T. Вершину v дерева T назовем *ветвящейся*, если среди ее сыновей есть по крайней мере два, из которых достижимы выделенные листья. В противном случае, т.е. когда только один сын вершины v имеет выделенных потомков в кроне дерева T, будем называть такую вершину *неветвящейся*.

Построим в дереве T путь $v_1, v_2,\ldots,v_p$ следующим образом.

1) Вершина v_1 есть корень;

2) Если вершина v_i построена, то при условии, что она неветвящаяся, вершину v_{i+1} определим как того сына вершины v_i, из которого достижимы отмеченные листья; если же вершина v_i ветвящаяся, то выберем вершину v_{i+1} как такого сына вершины v_i, из которого достижимо наибольшее число выделенных листьев.

   (Если таких «плодовитых» сыновей несколько, можно условиться выбирать среди них, скажем, самого правого.)

3) Если v_i лист, то это последняя вершина пути.

Пусть q_i означает число выделенных листьев, достижимых из вершины v_i построенного выше пути, а r_i означает число ветвящихся вершин – подлинных предков вершины v_i. Докажем, что имеет место неравенство

$$q_i \geq l^{2m+3-r_i}.$$

Индукция по i: при $i = 1$ вершина v_1 - корень, $r_1 = 0$, и $q_1 \geq l^{2m+3}$ - очевидно, так как из корня достижимы все листья и, в частности, выделенные. Пусть доказываемое соотношение верно для любого $n \leq i < p$. Тогда если вершина v_i неветвящаяся, то выполняется равенство

$$q_{i+1} = q_i \geq l^{2m+3-r_i},$$

причем $r_{i+1} = r_i$. Если же вершина v_i ветвящаяся, то наименьшее число выделенных листьев, достижимых из вершины v_{i+1}, получается, если из всех сыновей вершины v_i, «братьев» вершины v_{i+1}, наибольшее число которых равно l, достижимо одинаковое число выделенных листьев – $(1/l)$-я часть всех выделенных листьев, достижимых из отца – вершины v_i. Тогда

$$q_{i+1} \geq l^{2m+3-r_i} / l = l^{2m+3-(r_i+1)},$$

причем $r_{i+1} = r_i + 1$.

Логарифмируя неравенство

$$q_i \geq l^{2m+3-r_i},$$

получим:

$$\log_l q_i \geq 2m + 3 - r_i,$$

откуда

$$r_i \geq 2m + 3 - \log_l q_i.$$

Для листа v_p имеем $q_p = 1$ и $r_p \geq 2m + 3$. Это значит, что лист v_p имеет по меньшей мере $2m + 3$ подлинных предков, являющихся ветвящимися вершинами, и поэтому $p > 2m + 3$. Таким образом, построенный выше путь $v_1, v_2, \ldots, v_p$ имеет длину, не меньшую $2m + 3$.

Пусть R - нетерминал, служащий меткой такой вершины пути $v_1, v_2, \ldots, v_p$, которая является $(2m+3)$-ей от конца пути ветвящейся вершиной, т.е. если это вершина v_k, то

$v_k, v_{k+s}, \ldots, v_p$, $s \geq 1$ - последние $2m+3$ ветвящиеся вершины пути $v_1, v_2, \ldots, v_p$ (см. рис. 3).

Ветвящуюся вершину пути $v_1, v_2, \ldots, v_p$ назовем *левой (правой) ветвящейся* вершиной, если ее сын, не принадлежащий указанному пути, имеет достижимый из него выделенный лист левее (правее) листа v_p.

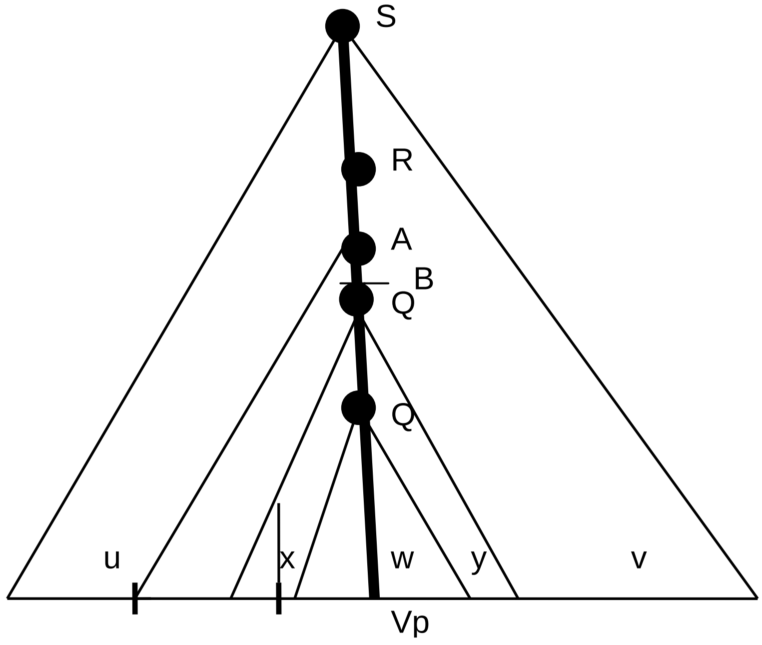

Рис. 3

Представим число $2m+3$ в виде: $2m+3 = (m + 2) + (m + 1)$. Может оказаться, что среди ветвящихся вершин пути $v_1, v_2,\ldots,v_p$ преобладают левые или, соответственно, правые ветвящиеся вершины. Разберем случай, когда преобладают левые (второй случай анализируется аналогично), т.е. таких вершин не менее $m + 2$. Выделим в пути $v_1, v_2,\ldots,v_p$ последние $m + 2$ левых ветвящихся вершины и предположим, что последняя от конца такая вершина помечена нетерминалом A. Тогда, как в доказательстве леммы о разрастании, заключаем, что на пути из A в v_p некая вершина (также левая ветвящаяся), помеченная Q повторяется хотя бы дважды.

При этом, поскольку вершина A есть $(m+2)$- я левая ветвящаяся вершина от конца пути, существует некая вершина, помеченная нетерминалом B, являющаяся $(m+1)$ -ой от конца левой ветвящейся вершиной. Тогда верхнюю вершину Q можно считать не совпадающей с A так как повторение нетерминала гарантировано и на пути от B. Выделяя максимальные поддеревья дерева T, с корневыми вершинами, помеченными Q, точно также как в доказательстве леммы о разрастании придем к представлению цепочки z в виде соединения пяти цепочек $z = uxwyv$ и возможности «накачки» цепочек x, y (условие (4)).

А именно, в силу выделения указанных выше максимальных поддеревьев, получаем возможность представить вывод цепочки z из аксиомы следующим образом:

$$S \vdash^* uQv \vdash^* uxQyv \vdash^* uxwyv.$$

(Здесь и далее символом $\vdash$ обозначено отношение непосредственной выводимости на множестве цепочек в объединенном алфавите грамматики (см. [1]), а символом $\vdash^*$ его рефлексивно-транзитивное замыкание.)

Это доказывает представление цепочки z в виде $z = uxwyv$.

Но тогда в рамках этого вывода можно сколько угодно раз повторить вывод из Q цепочки xQy, получив тем самым любую цепочку ux^nQy^nv ($n > 0$), а из нее вывести цепочку ux^nwy^nv. Но можно вывод из Q цепочки xQy вовсе опустить, перейдя сразу к выводу средней цепочки w из Q. Таким образом, цепочки x и y можно «накачать» (т.е. повторить их сколько угодно раз) или вовсе выбросить.

Условие (1) выполняется, так как существует сын вершины A, не принадлежащий пути $v_1, v_2,\ldots,v_p$, из которого достижим некоторый выделенный лист, метка которого входит в цепочку u, а из такого же свойства сына «верхней» вершины Q следует, что из него достижим лист, метка которого входит в цепочку x. Это справедливо в рассмотренном выше случае преобладания левых ветвящихся вершин в пути $v_1, v_2,\ldots,v_p$; в противоположном случае аналогичный вывод будет сделан для цепочек y, v.

Выделенной позицией цепочки w будет лист v_p (условие (2)).

Вершина R является $(2m+3)$-ей от конца пути $v_1, v_2,\ldots,v_p$ ветвящейся вершиной. Следовательно, из нее достижимо не более l^{2m+3} выделенных листьев, так как каждая ветвящаяся вершина имеет не более l ветвящихся потомков.. Отсюда подавно цепочка xwy содержит не более $l^{2m+3} = k_L$ выделенных позиций. Таким образом, выполняется условие (3).

Ясно, что сформулированная выше лемма о разрастании для кс-языков является частным случаем леммы Огдена, когда все вхождения букв считаются выделенными.

Рассмотрим пример применения леммы.

Докажем, что язык $L_5 = \{a^mb^nc^l \mid m, n, l$ попарно различны$\}$ не является кс-языком.

Предположим, что данный язык контекстно-свободный. Возьмем цепочку $z=a^k b^{k+(k-1)!} c^{k+k!}$, где k - константа из леммы Огдена, выделив в ней все вхождения символа a. Тогда при представлении цепочки z в виде $uxwyv$ цепочка w обязательно «зацепит» хотя бы один символ a(см. рис. 4); следовательно, цепочка x состоит только из символов a (как и цепочка u). А именно,

$$x = a^p, 1 \le p \le k-1.$$

Тогда, если цепочка w содержит и другие символы, кроме a, цепочка y может входить либо в «зону» символов b (целиком), либо в «зону» символов c (целиком), так как расположение «накачиваемых» цепочек на «стыках зон», очевидно, невозможно. В первом случае «кратность» α накачки цепочки x, которая уравняет числа символов a и c, определяется из соотношения:

$$k + \alpha p = k + k!, \text{ т.е. } \alpha = \tfrac{k!}{p}.$$

Во втором случае $(k\text{-}1)!/p)$ - кратная накачка цепочки x уравняет числа вхождений символов a и b.

Не исключено, наконец, что обе накачиваемые цепочки расположены в «зоне» символов a. Но тогда одним из указанных выше способов накачки можно уравнять числа либо символов a и b, либо a и c.

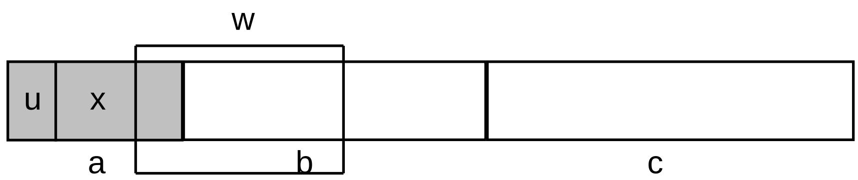

Рис. 4

Заметим, что возможность выделения символов существенно упрощает анализ данного языка, так как позволяет считать, что цепочка x может расположиться единственным способом. Иначе, т.е. при использовании леммы о разрастании для кс-языков, решение задачи было бы, по меньшей мере, сильно затруднено.

3. Лемма о разрастании для линейных языков

Теорема 4 (*Лемма о разрастании для линейных языков*) Для каждого линейного языка L существует константа k такая, что каждое слово z языка L, длина которого больше k, представимо в виде $z = uvwxy$, причем длина $|uvxy|$ не превосходит k, , $|vx| > 0$, и для каждого неотрицательного целого n слово $z_n = uv^nwx^ny$ принадлежит языку L.

Доказательство. Рассмотрим линейную грамматику $G=(V, N, S, P)$, порождающую язык L. Без ограничения общности считаем, что эта грамматика задана в приведенной форме. Рассмотрим множество D всех выводов в грамматике G таких, что их длина и, в силу линейности грамматики, равная ей высота дерева вывода, не превосходит числа $|N| + 1$. Положим

$$k = \max_{d \in D} |\alpha_d|,$$

где α_d - последняя цепочка вывода d.

Пусть $S \vdash^* z$, и $|z| > k$. Тогда длина вывода цепочки z из аксиомы S будет больше, чем $|N| + 1$. Пусть теперь R - нетерминал грамматики G такой, что $S \vdash^* u_1Ru_2$ и $u_1Ru_2 \vdash^* u_1w' y = z$, причем длина вывода цепочки u_1Ru_2 из аксиомы *равна* $|N| + 1$ (такой, единственный в силу линейности грамматики, нетерминал найдется, так как длина вывода терминальной цепочки z строго больше числа $|N| + 1$). Это значит, что какой-то из нетерминалов Q, фигурирующих в этом выводе, повторяется, и существуют, таким образом, выводы:

$$S \vdash^* uQy \vdash^* uvQxy \vdash^* uvw_1Rw_2\, xy \vdash^* uvw_1w'\, w_2\, xy = z,$$

т.е.

$$uvw_1 = u_1, w_2xy = u_2, w = w_1w'w_2 .$$

Итак, $S \vdash^* uvwxy$. Повторяя многократно вывод $Q \vdash^* vQx$, или выбрасывая его вовсе, получим, что для каждого неотрицательного целого n слово $z_n = uv^nwx^ny$ принадлежит языку L. Далее: длина вывода цепочки $uvQxy$ из цепочки из аксиомы не превышает $|N| + 1$, откуда $|uvQxy| \le k$ и подавно $|uvxy| \le k$. То, что $|vy| > 0$, следует из приведенности грамматики G .

Рассмотрим пример.

Докажем, что язык правильных скобочных структур и язык, состоящий из всех слов в алфавите $\{a, b\}$ таких, что число символов a и символов b в них совпадает, не являются линейными.

Положим, что язык правильных скобочных структур линеен. Обозначая через a открывающую, а через b - закрывающую скобку, рассмотрим цепочку

$z = a^{2k}b^{2k}a^{4k}b^{4k}$, где k - константа из леммы о разрастании для линейных языков. Так как $|z| = 12k$, то по лемме $z = uvwxy$, причем $|w| \geq 11k$.

Следовательно, возможны два случая:

1) $w = a^s b^{2k} a^{4k} b^{4k}$, где $s \geq k$ и

2) $w = a^{2k} b^{2k} a^{4k} b^p$, где $p \geq 3k$.

В первом случае

$$v = a^r, x = y = \lambda \text{, где } 0 < r \leq s;$$

во втором - $x = b^q, u = v = \lambda, 0 < q \leq p$.

Ясно, что накачка невозможна ни в том, ни в другом случае.

Та же цепочка может быть взята и для анализа второго языка.

4. Об одном примере использования леммы о разрастании

Здесь мы рассматриваем пример, в котором лемма о разрастании для КС-языков используется «позитивно», т.е. с ее помощью мы доказываем принадлежность некоторого языка к определенному классу.

Теорема 5. Всякий КС-язык в однобуквенном алфавите регулярен.

Для любого языка L в алфавите $\{a\}$ имеет место биекция

$$f : L \to N_L \subseteq N$$

такая, что $f(x) = |x|$. Для КС-языка L в силу леммы о разрастании на множестве N_L всякое число, строго большее фиксированной константы k_L, является членом некоторой арифметической прогрессии. Разность любой такой прогрессии (совокупная длина накачиваемых цепочек) ограничена сверху той же константой k_L. Значит, множество всех указанных прогрессий конечно. Следовательно, каждое число множества N_L является членом одной из конечного множества арифметических прогрессий. Итак, множество N_L представляется в виде конечного объединения: множеств членов прогрессий и конечного множества чисел, не превосходящих k_L. Пусть d - разность какой-либо из упомянутых арифметических прогрессий, и пусть n_0 - ее начальный член. Тогда соответствующий язык в алфавите $\{a\}$ порождается следующей праволинейной грамматикой:

$$\begin{array}{l} S \to a^{n_0} \mid a^{n_0}T \\ T \to a^d T \mid a^d \end{array},$$

для которой может быть построена эквивалентная регулярная грамматика. Язык же, который состоит из всех слов языка L, длина которых не больше k_L, конечен и потому регулярен. Следовательно, язык L может быть представлен как конечное объединение регулярных языков и является регулярным.

Заключение

Основными результатами статьи являются подробно изложенные доказательства трех теорем: леммы Огдена о КС-языках, леммы о разрастании для линейных языков и утверждения о регулярности любого КС-языка в однобуквенном алфавите. При этом доказательства второй и третьей теоремы принадлежат автору статьи.

Кроме того, в рамках изложенной методики рассмотрения всех лемм о разрастании, одного из ключевых разделов теории формальных языков, особый акцент сделан на анализе нетривиальных примеров, обычно не рассматриваемых в известных руководствах. К числу таких примеров принадлежат примеры языков, в которых числа вхождений символов удовлетворяют некоторым неравенствам, языков, удовлетворяющих леммам о разрастании, но не принадлежащих соответствующему классу языков. На одном важном примере обсуждается метод анализа, названный «методом факториальной накачки» и основанный на лемме Огдена.

В целом, предлагаемая система изложения позволяет глубже оценить возможности лемм о разрастании в плане анализа языков на предмет их принадлежности или непринадлежности к тому или иному классу языков.

Список литературы

1. *А.И. Белоусов, С.Б. Ткачев*. Дискретная математика. – 5-е изд. - М.: Изд-во МГТУ им. Н.Э. Баумана, 2015. – 744 с.
2. *Дж. Хопкрофт, Р. Мотвани, Дж. Ульман*. Введение в теорию автоматов, языков и вычислений, 2-е изд.. : Пер. с англ. – М.: Издательский дом «Вильямс», 2002. – 528 с.
3. *А. Ахо, Дж. Ульман*. Теория синтаксического анализа, перевода и компиляции: Пер. с англ: В 2 т. Т. 1. М.: Мир, 1978.- 612 с.
4. *А.Е. Пентус, М.Р. Пентус*. Математическая теория формальных языков. – М.: Интернет-университет информационных технологий; БИНОМ, Лаборатория знаний, 2006. – 247 с.