

УДК 519.6

©2009. Е.А. Пряничникова

ХАРАКТЕРИЗАЦИЯ ЯЗЫКОВ, ПРЕДСТАВИМЫХ В ГРАФАХ С ОТМЕЧЕННЫМИ ВЕРШИНАМИ

Для языков, представимых графами с отмеченными вершинами, доказывается теорема, аналогичная теореме Майхилла-Нерода для языков, распознаваемых конечными автоматами. На основе доказательства теоремы показано, что любому языку, порождаемому графом с отмеченными вершинами, соответствует единственный с точностью до изоморфизма полный детерминированный граф с минимальным числом вершин и единственная система левосторонних уравнений с минимальным числом уравнений, из решений которой может быть получено регулярное выражение, описывающее этот язык.

Введение. В настоящее время актуальны задачи, связанные с анализом графов с отмеченными вершинами [1, 2]. В работе [3] введена алгебра языков, представимых такими графами, для которой доказывается теорема, аналогичная теореме Клини для конечных автоматов: любому регулярному выражению этой алгебры соответствует такой граф с отмеченными вершинами, что язык, представимый этим графом, совпадает с языком, описываемым регулярным выражением. В связи с этим представляет интерес дальнейшее исследование свойств рассматриваемой алгебры с целью получения аналогов остальных важных результатов, известных для алгебры Клини.

Основная цель данной работы – доказать теорему, аналогичную теореме Майхилла-Нерода, и показать, что любому языку, порождаемому графом с отмеченными вершинами, соответствует единственный с точностью до изоморфизма полный детерминированный граф с минимальным числом вершин и единственная система левосторонних уравнений с минимальным числом уравнений, из решений которой может быть получено регулярное выражение, описывающее этот язык.

Работа имеет следующую структуру. В разделе 1 вводятся основные определения и обозначения. В разделе 2 доказывается, что классы языков, определяемых детерминированными и недетерминированными графами с отмеченными вершинами совпадают. В разделе 3 доказывается теорема, аналогичная теореме Майхилла-Нерода для конечных автоматов, следствием из которой является существование канонического графа с отмеченными вершинами. Показано, что для любого языка, представимого графом с отмеченными вершинами, есть единственная система уравнений минимального порядка, из решений которой может быть получено каноническое регулярное выражение, описывающее этот язык.

1. Основные определения. Назовем графом с отмеченными вершинами четверку $G = (Q, E, X, \mu)$, где Q – конечное множество вершин, $|Q| = n$, $E \subseteq Q \times Q$ – множество дуг, X – конечный алфавит отметок вершин, $\mu : Q \rightarrow X$ – функция отметок вершин. Пусть $I \subseteq Q$ – множество начальных, а $F \subseteq Q$ – множество финальных вершин графа G .

Назовем граф G детерминированным, если для любой пары $(q_i, x_i) \in Q \times X$ найдется не больше одной вершины $q_j \in Q$, для которой выполняется, что $(q_i, q_j) \in E$ и $\mu(q_j) = x_i$. Если для любой пары $(q_i, x_i) \in Q \times X$ в графе G найдется не меньше одной вершины $q_j \in Q$, для которой выполняется, что $(q_i, q_j) \in E$ и $\mu(q_j) = x_i$, то граф G назовем полным.

Путем в графе G будем называть конечную последовательность вершин $l = q_1 q_2 \dots q_k$, где $(q_i, q_{i+1}) \in E$. Число k будем называть длиной, q_1 – начальной, а q_k – конечной вершиной пути l . Отметкой пути l будем называть последовательность отметок вершин $\mu(q_1)\mu(q_2)\dots\mu(q_k)$. Языком, порожденным вершиной q_i , назовем множество отметок всех путей в графе G , у которых начальной вершиной является вершина q_i , а конечная вершина принадлежит множеству финальных вершин. Множество отметок всех путей в графе G , начальные вершины которых принадлежат множеству I , а конечные – множеству F , назовем языком, порождаемым графом G , и обозначим $L(G)$.

Рассмотрим алгебру $\langle 2^{X^+}, \circ, \cup, *, \emptyset, X \rangle$ со следующими операциями на языках $L, R \in 2^{X^+}$.

- 1) $L \cup R = \{w | w \in L \text{ или } w \in R\}$;
- 2) $L \circ R = \{w_1 \circ w_2 | w_1 \in L \text{ и } w_2 \in R\}$, где операция $\circ$ на множестве слов в конечном алфавите X определяется следующим образом: для всех $w_1, w_2 \in X^+$ и всех $x, y \in X$

$$w_1 x \circ y w_2 = \begin{cases} w_1 x w_2, & \text{если } x = y; \\ \text{не определено,} & \text{в противном случае;} \end{cases}$$
- 3) $L^* = \bigcup_{i=0}^{\infty} L^i$, где $L^0 = X$; $L^{n+1} = L^n \circ L$ для всех $n \geq 0$;
- 4) $L^{\otimes} = L_{beg} \circ L^* \circ L_{end}$, где $L_{beg} = \{x | xw \in L, x \in X, w \in X^*\}$; $L_{end} = \{x | wx \in L, x \in X, w \in X^*\}$.

Регулярные выражения в алгебре $\langle 2^{X^+}, \circ, \cup, \otimes, \emptyset, X \rangle$ определим индуктивно:

- 1) $\emptyset$ является регулярным выражением;
- 2) x и xu являются регулярными выражениями для всех $x, y \in X$;
- 3) если p и q – регулярные выражения, то выражения $(p \circ q), (p \cup q), (p^{\otimes})$ также являются регулярными.

Язык, обозначаемый регулярным выражением R , будем обозначать $L(R)$. Два регулярных выражения будем называть эквивалентными, если совпадают языки, которые они обозначают.

2. Детерминизация графов с отмеченными вершинами. Основная цель данного раздела – показать, что для любого графа с отмеченными вершинами G , порождающего язык $L(G)$, существует единственный с точностью до изоморфизма минимальный полный детерминированный граф G' , порождающий такой же язык.

По аналогии с конечными автоматами, введем определение функции переходов для графов с отмеченными вершинами. Обозначим через X^2 все возможные пары символов алфавита X . Для недетерминированного графа $G = (Q, E, X, \mu)$ назовем функцией переходов такую функцию $\delta : Q \times X^2 \rightarrow 2^Q$, что

$$\delta(q_1, xy) = \{q_2 \mid (q_1, q_2) \in E, \mu(q_1) = x, \mu(q_2) = y\}$$

для любых $q_1 \in Q, xy \in X^2$.

Расширим функцию δ до $\hat{\delta} : 2^Q \times X^+ \rightarrow 2^Q$ следующим образом: для любых $q_1 \in Q, Q_1 \subseteq Q, x \in X, xy \in X^2, w \in X^+$

$$\begin{aligned}\hat{\delta}(q_1, x) &= \begin{cases} q_1, & \text{если } \mu(q_1) = x; \\ \emptyset, & \text{иначе,} \end{cases} \\ \hat{\delta}(q_1, xy) &= \{q_2 \mid (q_1, q_2) \in E, \mu(q_1) = x, \mu(q_2) = y\}, \\ \hat{\delta}(q_1, w \circ xy) &= \{q_2 \mid q_2 \in \delta(q_3, xy) q_3 \in \hat{\delta}(q_1, w)\}, \\ \hat{\delta}(Q_1, w) &= \bigcup_{q \in Q_1} \hat{\delta}(q, w).\end{aligned}$$

Таким образом, для любого слова w и некоторого подмножества вершин графа Q_1 функция $\hat{\delta}(Q_1, w)$ описывает множество вершин графа G , в которые можно попасть из вершин, входящих в множество Q_1 по путям, отметки которых совпадают со словом w . В том случае, когда $G = (Q, E, X, \mu)$ является полным детерминированным графом, функция переходов $\delta : Q \times X^2 \rightarrow Q$, определяется следующим образом: для любых $q_1 \in Q, xy \in X^2$

$$\delta(q_1, xy) = q_2 \mid (q_1, q_2) \in E, \mu(q_1) = x, \mu(q_2) = y.$$

Расширенная функция $\hat{\delta} : 2^Q \times X^+ \rightarrow Q$ в этом случае будет определена как

$$\begin{aligned}\hat{\delta}(q_1, x) &= \begin{cases} q_1, & \text{если } \mu(q_1) = x; \\ \emptyset, & \text{иначе,} \end{cases} \\ \hat{\delta}(q_1, xy) &= q_2 \mid (q_1, q_2) \in E, \mu(q_1) = x, \mu(q_2) = y, \\ \hat{\delta}(q_1, w \circ xy) &= \delta(\hat{\delta}(q_1, w)), \\ \hat{\delta}(Q_1, w) &= \bigcup_{q \in Q_1} \hat{\delta}(q, w),\end{aligned}$$

где $q_1 \in Q, Q_1 \subseteq Q, x \in X, xy \in X^2, w \in X^+$.

Теорема 1. Для любого недетерминированного графа G , порождающего язык $L(G)$, существует такой полный детерминированный граф G' , что $L(G) = L(G')$.

Доказательство. Для доказательства используем метод, аналогичный конструкции подмножеств для конечных автоматов [4]. Пусть $G = (Q, E, X, \mu)$ – недетерминированный граф, $\hat{\delta}$ – его функция переходов.

Пусть алфавит X состоит из m символов $\{x_1, \dots, x_m\}$. Обозначим через Q_i множество всех вершин графа G с меткой x_i .

Определим граф $G' = (Q', E', X, \mu')$ следующим образом. Пусть $Q' = \{q' \mid q' \in \bigcup_{x_i \in X} 2^{Q_i}\}$. Для графа G' множество начальных вершин $I' = \{\{q_i\} \mid q_i \in I\}$, множество финальных вершин $F' = \{Q_i' \mid Q_i' \cap F \neq \emptyset\}$. Таким образом, начальными состояниями графа G' будут такие подмножества вершин графа G , которые состоят из одной вершины, и эта вершина является начальной. Все подмножества, которые включают хотя бы одну финальную вершину, будут финальными вершинами графа G' . Функцию переходов графа G' определим, как $\hat{\delta}'(Q_1', w) = Q_2'$, где $Q_2' = \hat{\delta}(Q_1', w)$.

Индукцией по длине слова w покажем, что $\hat{\delta}'(I', w) = \hat{\delta}(I, w)$. Если слово w состоит из одного символа, то $w = x, x \in X$ и $\hat{\delta}'(I', x) = \hat{\delta}(I, w) = \{q \mid \mu(q) = x, q \in I\}$. Пусть гипотеза верна для слов w длины k и меньше. Тогда для слова $w \circ xy$ длины $k + 1$

$$\hat{\delta}'(I', w \circ xy) = \bigcup_{\{q_i\} \in I'} \hat{\delta}'(\delta'(\{q_i\}, w), xy),$$

$$\hat{\delta}(I, w \circ xy) = \bigcup_{q_i \in I} \hat{\delta}(\delta(q_i, w), xy),$$

откуда по определению графа G'

$$\hat{\delta}'(I', w \circ xy) = \hat{\delta}(I, w \circ xy).$$

Таким образом, множество вершин графа G , в которое можно попасть из начальных вершин по любому слову w , совпадает с множеством, связанным с теми вершинами графа G' , в которые мы можем попасть по слову w из начальных вершин этого графа. Слово w принадлежит языку $L(G)$ в том случае, когда среди вершин, в которые мы попадаем по этому слову из начальных вершин, есть финальные. По определению графа G' , в этом случае слово w будет принадлежать и языку, порождаемому графом G' . Следовательно, $L(G) = L(G')$, что и требовалось доказать. $\square$

Нужно отметить, что есть существенное различие между детерминированными графами с отмеченными вершинами и детерминированными автоматами. У полного детерминированного автомата есть только одно начальное состояние и только одно "пустое" состояние. У полного детерминированного графа с отмеченными вершинами начальных и "пустых" вершин будет столько, сколько символов в алфавите. Различны и оценки количества состояний после детерминизации. Если у исходного недетерминированного автомата количество состояний было равно n , то при применении алгоритма конструкции подмножеств для его детерминизации у полученного автомата в худшем случае будет 2^n состояний. При применении алгоритма детерминизации графов с отмеченными вершинами, основанного на теореме 1, количество вершин полного детерминированного графа, полученного в результате, в худшем случае будет равно $|Q'| = \sum 2^{|Q_i|}$.

В то же время и у полного детерминированного графа и у полного детерминированного автомата есть важное общее свойство, на наличии которого основано доказательство теоремы Майхилла-Нерода: для любого слова из X^+ в них всегда есть единственный путь, ведущий из начальной вершины, отметка которого совпадает с этим словом.

3. Аналог теоремы Майхилла-Нерода для графов с отмеченными вершинами. Используя принятую в [4] терминологию и систему обозначений, введем понятие производной языка $L \subseteq X^+$, представимого конечным ориентированным графом с отмеченными вершинами, по слову $w \in X^+$:

$$D_w L = \{y \mid w \circ y \in L\}. \quad (1)$$

Для языка $L \subseteq X^+$ определим отношение $\rho_L \subseteq X^+ \times X^+$ следующим образом: для всех $u, v \in X^+$ $u\rho_L v$ тогда и только тогда, когда $D_u L = D_v L$ и слова u и v заканчиваются одним и тем же символом $x \in X$. Легко проверить, что это отношение рефлексивно, симметрично, транзитивно и является право-инвариантным.

Следующая теорема является аналогом теоремы Майхилла-Нерода для конечных автоматов.

Теорема 2. *Язык $L \subseteq X^+$ порождается конечным графом с отмеченными вершинами тогда и только тогда, когда отношение ρ_L разбивает X^+ на конечное число классов эквивалентности.*

Доказательство. Покажем, что если язык L порождается конечным графом с отмеченными вершинами, то число классов конечно.

Пусть язык L порождается полным детерминированным графом $G = (Q, E, X, \mu)$. Для любого слова $w \in X^+$ в графе G есть единственный путь, ведущий из начальной вершины, отметка которого совпадает со словом w . Пусть этот путь заканчивается в некоторой вершине $q \in Q$.

Для любого графа G можно определить отношение эквивалентности $\pi_G \subseteq X^+ \times X^+$ таким образом, что два слова u и v попадают в один класс тогда, когда идущие из начальных вершин пути в графе G , отметки которых совпадают с этими словами, заканчиваются в одной и той же вершине. Так как в графе конечное число вершин, то число классов эквивалентности, на которые множество X^+ разбивает отношение π_G , также будет конечным.

Если для двух слов u и v пути из начальных вершин заканчиваются в одной и той же вершине $q \in Q$, то $D_u L = D_v L$. Таким образом, все слова $w \in X^+$, которые попадают в один класс эквивалентности по отношению π_G , принадлежат к одному из классов эквивалентности, на которые множество X^+ разбивает отношение ρ_L . Отсюда следует, что число классов эквивалентности у отношения ρ_L тоже конечно.

Для доказательства того, что в том случае, когда отношение ρ_L разбивает X^+ на конечное число классов n , язык L порождается некоторым графом G с отмеченными вершинами, построим граф $G = (Q, E, X, \mu)$ следующим образом.

В графе G будет n состояний, каждое из которых соответствует одному из классов эквивалентности отношения ρ_L . По определению этого отношения все слова, входящие в один класс эквивалентности, заканчиваются на один и тот же символ $x \in X$. Этот символ x будет отметкой соответствующего классу состояния. Отношение ρ_L право-инвариантно по операции $\circ$, то есть для всех $x, y \in X^+$, если $x\rho_L y$, то $x \circ z\rho_L y \circ z$, следовательно, для всех слов w_i , которые принадлежат некоторому классу и заканчиваются на один и тот же символ $x \in X$, все слова $w_i \circ xy$, где $x, y \in X$, тоже принадлежат одному классу. Это свойство дает возможность определить множество дуг E в графе G следующим образом. Для каждого состояния q_i , для которого $\mu(q_i) = x$ дуга $(q_i, q_j) \in E$ в том случае, когда $\mu(q_j) = y$ и для всех слов w_i , входящих в класс, соответствующий состоянию q_i , все слова $w_i \circ xy$ принадлежат классу, соответствующему состоянию q_j . Начальными в графе будут те состояния q , классы эквивалентности слов для которых содержат символы алфавита X как отдельные слова. Конечными будут те состояния, классы которых содержат слова,

входящие в язык L . Индукцией по длине слова w легко показать, что в построенном таким образом полном детерминированном графе путь с отметкой w , ведущий из начального состояния в конечное, есть тогда и только тогда, когда $w \in L$, следовательно, этот граф порождает язык L . $\square$

Теорема 3. *Для любого языка L , порождаемого графом с отмеченными вершинами, граф, построенный в доказательстве теоремы 2, является единственным с точностью до изоморфизма минимальным по количеству вершин полным детерминированным графом, порождающим язык L .*

Доказательство. Пусть существует полный детерминированный граф $G' = (Q', E', X, \mu')$ такой, что $L(G') = L(G)$ и число состояний графа G' меньше, чем n . Тогда в таком графе есть такая вершина, в которой заканчиваются пути из начальных вершин с отметками, принадлежащими разным классам эквивалентности отношения ρ_L $D_u L \neq D_v L$, что невозможно. Предположим, что число вершин у графов G и G' одинаково. Каждая из вершин графа G соответствует одному из классов эквивалентности. В графе G' отметки всех путей, ведущих из начального состояния в состояние q' , входят в один из классов эквивалентности отношения ρ_L и образуют один из классов эквивалентности отношения π_G . Так как число классов у обоих отношений одинаково и отношения определены на одном и том же множестве, то есть взаимно однозначное соответствие между множествами вершин Q и Q' , следовательно, графы G и G' изоморфны, что и требовалось доказать. $\square$

Теорему, аналогичную теореме Майхилла-Нерода, можно доказать иначе, используя для доказательства не свойства путей в графах с отмеченными вершинами, а свойства регулярных выражений, описывающих языки, порождаемые графами.

Теорема 4. *Пусть отношение $\rho_L \subseteq X^+ \times X^+$ для языка $L \subseteq X^+$ определяется следующим образом: для всех $u, v \in X^+$ $u \rho_L v$ тогда и только тогда, когда $D_u L = D_v L$ и слова u и v заканчиваются одним и тем же символом $x \in X$. Язык L можно представить регулярным выражением алгебры $\langle 2^{X^+}, \circ, \cup, \otimes, \emptyset, X \rangle$ тогда и только тогда, когда отношение ρ_L разбивает множество X^+ на конечное число классов эквивалентности.*

Доказательство. Для регулярных выражений алгебры $\langle 2^{X^+}, \circ, \cup, \otimes, \emptyset, X \rangle$ по аналогии с [5] введем понятие производной.

Рекурсивно определим производную $D_w R$ регулярного выражения R по слову $w \in X^+$ следующим образом.

Для любого $x \in X$, $xy \in X^2$, и регулярных выражений R и Q

$$D_x \emptyset = \emptyset; \quad (2)$$

$$D_x x = x; D_x xy = xy; \quad (3)$$

$$D_x y = \emptyset, D_x yz = \emptyset, y \neq x, yz \in X^2; \quad (4)$$

$$D_{xy} xy = y; \quad (5)$$

$$D_{xy} zt = \emptyset, zt \in X^2, zt \neq xy; \quad (6)$$

$$D_{xy}(R \cup Q) = D_{xy}(R) \cup D_{xy}(Q); \quad (7)$$

$$D_{xy}(R \circ Q) = D_{xy}R \circ Q \cup \varepsilon(R) \circ D_{xy}Q; \quad (8)$$

$$D_{xy}(R^{\otimes}) = D_{xy}(R) \cup D_{xy}(\alpha) \circ R^{\otimes}; \quad (9)$$

$$D_{xy \circ yz}R = D_{yz}(D_{xy}R). \quad (10)$$

Отображение ε определено таким образом, что $\varepsilon(R)$ – это объединение всех букв алфавита X , которые входят в язык, представляемый регулярным выражением R , как отдельные слова. Для любого регулярного выражения R определим отношение $\rho_R \subseteq X^+ \times X^+$ следующим образом: для всех $u, v \in X^+$ $u \rho_R v$ тогда и только тогда, когда $D_u R = D_v R$ и слова u и v заканчиваются одним и тем же символом $x \in X$.

Лемма 1. *Для любого регулярного выражения R среди регулярных выражений, которые можно получить, вычисляя производные по всем словам $w \in X^+$ пользуясь правилами (2–10), есть только конечное число различных формул.*

Доказательство леммы. Для доказательства воспользуемся методом индукции по структуре формулы R . Если $R = X$, или $R = \emptyset$, или $R = a$, где $a \in X$, то конечность числа различных производных следует из определения D_x .

Предположим, что $R = R_1 \cup R_2$, где R_1 и R_2 формулы, число различных производных которых конечно и равно n_1 и n_2 соответственно. Тогда из определения производных следует, что число различных производных формулы R $n = n_1 + n_2$, то есть тоже конечно. Аналогичными рассуждениями можно показать, что число производных формулы R будет конечным и в тех случаях, когда $R = R_1 \circ R_2$ и $R = R_1^{\otimes}$, что и требовалось доказать. $\square$

Из леммы 1 следует, что отношение ρ_R делит множество X^+ на конечное число классов эквивалентности.

Покажем, что, если слова $u \in X^+$ и $v \in X^+$ входят в один и тот же класс эквивалентности по отношению ρ_R , то они входят в один и тот же класс эквивалентности по отношению ρ_L .

Лемма 2. *Пусть регулярное выражение R описывает язык $L(R)$. Тогда для любого слова $w \in X^+$ производная регулярного выражения R по слову w описывает язык, совпадающий с производной языка $L(R)$ по слову w : $L(D_w R) = D_w L(R)$.*

Доказательство леммы. По определению регулярных выражений, возможны следующие случаи:

$$R = \emptyset; R = x, x \in X; R = xy, x, y \in X; R = R_1 \cup R_2, R = R_1 \circ R_2, R = R_1^{\otimes}.$$

Докажем лемму по индукции. Если $R = x, x \in X$, то тогда $L(R) = \{x\}$. По определению производных регулярных выражений, у формулы R в этом случае есть только 2 различные производные:

$$D_x x = x,$$

$$D_w x = \emptyset, \text{ если } w \in X^+ \text{ и } w \neq x.$$

У языка $L(R) = \{x\}$ есть тоже только 2 различные производные:

$$D_x L(R) = \{x\},$$

$$D_w L(R) = \emptyset, \text{ если } w \in X^+ \text{ и } w \neq x.$$

Таким образом, в обоих случаях $L(D_w R) = D_w L(R)$.

Пусть $R = xy, x, y \in X$. В этом случае у формулы R есть три различные производные:

$$D_x xy = xy, D_{xy} xy = y, D_w x = \emptyset, \text{ если } w \in X^+ \text{ и } w \neq xy.$$

У языка $L(R) = \{xy\}$ есть тоже только 2 различные производные:

$$D_x \{xy\} = \{xy\}, D_{xy} \{xy\} = \{y\}, D_w \{xy\} = \emptyset, \text{ если } w \in X^+ \text{ и } w \neq xy, \text{ таким образом, в этом случае утверждение леммы также выполняется и } L(D_w R) = D_w L(R).$$

Если $R = \emptyset$, то единственная производная регулярного выражения R $D_w R = \emptyset$ для всех $w \in X^+$, а единственная производная языка $L(R)$ тоже равна $\emptyset$. Пусть $R = R_1 \cup R_2$, причем известно, что для всех $w \in X^+$ $L(D_w R_1) = D_w L(R_1)$, $L(D_w R_2) = D_w L(R_2)$. По определению производных регулярных выражений $D_w(R_1 \cup R_2) = D_w R_1 \cup D_w R_2$, откуда $L(D_w R) = L(D_w(R_1 \cup R_2)) = L(D_w R_1 \cup D_w R_2) = L(D_w R_1) \cup L(D_w R_2)$.

По определению производных языков

$$D_w L(R_1) = \{y | w \circ y \in L(R_1)\};$$

$$D_w L(R_2) = \{y | w \circ y \in L(R_2)\}.$$

Тогда $D_w L(R_1) \cup D_w L(R_2) = \{y | w \circ y \in L(R_1) \cup L(R_2)\} = \{y | w \circ y \in L(R)\} = D_w L(R)$, значит $L(D_w R) = D_w L(R)$, что и требовалось доказать. Пусть $R = R_1 \circ R_2$ и для всех $w \in X^+$ $L(D_w R_1) = D_w L(R_1)$, $L(D_w R_2) = D_w L(R_2)$. Тогда $L(D_w R) = L(D_w(R_1 \circ R_2)) = L(D_w R_1) \circ L(R_2) \cup L(\varepsilon(R_1)) \circ L(D_w R_2) = D_w L(R_1) \circ L(R_2) \cup L(\varepsilon(R_1)) \circ D_w L(R_2)$. $D_w L(R) = \{y | y \circ w \in L(R_1) \circ L(R_2)\} = \{u \circ v | w \circ u \in L(R_1), v \in L(R_2)\} \cup \{u \circ w \circ v | u \in L(R_1), u \in X, w \circ v \in L(R_2)\} = \{u | w \circ u \in L(R_1)\} \circ \{v | v \in L(R_2)\} \cup \{u | u \in L(R_1), u \in X\} \circ \{v | w \circ v \in L(R_2)\} = D_w L(R_1) \circ L(R_2) \cup L(\varepsilon(R_1)) \circ D_w L(R_2) = L(D_w R)$, то есть $L(D_w R) = D_w L(R)$, что и требовалось доказать. В третьем случае, когда $L(D_w R) = L(D_w(R_1^*)) = L(D_w R_1^0) \cup L(D_w R_1) \cup L(D_w R_1) \circ L(R_1^*)$. $D_w L(R) = D_w L(R_1^*) = \{u | w \circ u \in L(R_1^*)\} = \{u | w \circ u \in L(R_1^0)\} \cup \{u | w \circ u \in L(R_1)\} \cup \{u | w \circ u \in L(R_1) \circ L(R_1^*)\} = L(D_w R)$, следовательно, во всех случаях $L(D_w R) = D_w L(R)$, что и требовалось доказать. $\square$

Из леммы 2 следует, что для любого регулярного выражения R и слов $u, v \in X^+$ из того, что $u \rho_R v$ или $D_u R = D_v R$ следует, что $L(D_u R) = L(D_v R)$, откуда $D_u L(R) = D_v L(R)$, что означает, что $u \rho_{L(R)} v$, то есть слова u и v находятся в одном классе по отношению $\rho_{L(R)}$. Таким образом, если у отношения ρ_R конечное число классов, то конечно и число классов у отношения $\rho_{L(R)}$, следовательно, доказана первая часть теоремы: если язык L может быть представлен регулярным выражением, то отношение ρ_L разбивает множество X^+ на конечное число классов эквивалентности.

Для доказательства второй части теоремы построим регулярное выражение R , описывающее язык $L(R)$, если известны классы, на которые отношение ρ_L разбивает множество X^+ . Внутри каждого из классов слова можно упорядочить по длине, а слова одной длины – в лексикографическом порядке, и каждому классу сопоставить минимальное слово в этом классе. Эти минимальные слова можно в свою очередь упорядочить таким же образом, одновременно упорядочивая и классы. Так как каждому классу соответствует одна из производных регулярного выражения R , то различные производные тоже можно упорядочить так, чтобы производной $D_i R$

соответствовал класс Y_i и минимальное слово в этом классе s_i .

По свойствам производных, каждую производную $D_i R$ можно записать в виде

$$D_i R = \bigcup_{ab \in X^2} ab \circ D_j R \cup \varepsilon(D_i R), \quad (11)$$

где $D_j R$ – это такая из различных производных, что слово $s_i \circ ab$ принадлежит классу Y_j .

В работе [3] показано, что система вида (11) имеет единственное решение. Из определения производных следует, что исходное регулярное выражение R можно записать в виде:

$$R' = \bigcup_{i=1}^n s_i \circ D_i R, \quad (12)$$

причем $L(R) = L(R')$. Таким образом, язык L можно представить регулярным выражением алгебры $\langle 2^{X^+}, \circ, \cup, \otimes, \emptyset, X \rangle$ тогда и только тогда, когда отношение ρ_L разбивает множество X^+ на конечное число классов эквивалентности, что и требовалось доказать. $\square$

Из доказательства теоремы 4 следует, что для любого языка L , представимого в графах с отмеченными вершинами, существует единственное регулярное выражение, построенное по формуле (12) из решений системы (11) порядка n , где n – число классов, на которые отношение ρ_L разбивает множество X^+ , причем для любого языка L минимальный порядок системы, из решений которой можно построить регулярное выражение, описывающее язык L , равен числу классов n .

Заключение. Для языков, представимых графами с отмеченными вершинами, доказана теорема, аналогичная теореме Майхилла-Нерода для языков, распознаваемых конечными автоматами. Показано, что любому языку, порождаемому графом с отмеченными вершинами, соответствует единственный с точностью до изоморфизма полный детерминированный граф с минимальным числом вершин, порождающий этот язык, и единственная система леволлинейных уравнений с минимальным числом уравнений, из решений которой может быть получено регулярное выражение, описывающее этот язык.

Автор выражает благодарность И.С.Грунскому за постановку задачи и полезные замечания.

1. Капитонова Ю.В., Летичевский А.А. Математическая теория проектирования вычислительных систем. – М.: Наука, 1988. – 296с.
2. Dudek G., Jenkin M., Milios E., Wilkes D. Map validation and robot self-location in a graph-like world // Robotics and autonomous systems, vol.22(2), November 1997. – P.159-178.
3. Грунский И.С., Пряничникова Е.А. Об алгебре языков, представимых графами с отмеченными вершинами // Труды ИПММ НАНУ, 2009. – Т.18. – С.37-46.
4. Yu S. Regular Languages // Handbook of Formal Languages, pps.41-110, edited by G. Rozenberg and A. Salomaa, Springer, 1997.
5. Brzozowski J.A. Derivatives of regular expressions // J. Assoc. Comput. Mach., 11: 481–494, 1964.