Теорема Майхилла-Нероуда (*конспект лекции*)

определяет необходимое и достаточное условия регулярности языка.

Она также позволяет доказать, что язык нерегулярен

Определение 1. Будем говорить, что слова $p,\ q \in \Sigma^*$ *различимы* словом $r \in \Sigma^*$ относительно языка $L \subseteq \Sigma^*$, если $pr \in L$, $qr \notin L$ или $pr \notin L$, $qr \in L$. Если для p и q различающих слов не существует, то будем говорить, что слова p и q *неразличимы* относительно языка L и писать $p \sim q(L)$ или $p \sim q$, когда язык L фиксирован. Таким образом,

$$p \sim q(L) \Leftrightarrow \forall\, r \in \Sigma^*\ (pr \in L \ \Leftrightarrow\ qr \in L).$$

Лемма 1. Отношение неразличимости слов относительно языка рефлексивно, симметрично и транзитивно, то есть является отношением эквивалентности.

Доказательство. Первые два свойства очевидны, поясним транзитивность. Пусть $p \sim q$, $q \sim w$ и r – произвольное слово. Транзитивность следует из соотношений $pr \in L \ \Leftrightarrow\ qr \in L \ \Leftrightarrow\ wr \in L$. ♦

Класс эквивалентности – это множество слов, для которых правые контексты совпадают, т.е. число множеств правых контекстов равно числу классов эквивалентности. Обозначим $[p]$, $p \in \Sigma^*$, класс эквивалентности, которому принадлежит p.

Множества правых контекстов

Определение 2. Пусть $L \subseteq \Sigma^*$ и $y \in \Sigma^*$. Тогда множество правых контекстов слова y относительно языка L определяется так:

$$C^{(r)}{}_L(y) \cong \{z \in \Sigma^* \mid yz \in L\}.$$

Пример 1. Пусть $\Sigma = \{a, b\}$ и $L = \{a^n b a^n \mid n \geq 0\}$. Тогда

1) $C^{(r)}{}_L(a^i) = \{a^k b a^{k+i} \mid k \geq 0\}$, число вхождений b в слово равно 0; для разных i множества разные, число таких множеств счётно;

$C^{(r)}{}_L(\varepsilon) = \{aba, a^2ba^2, \ldots\}$

$C^{(r)}{}_L(a) = \{ba, aba^2, a^2ba^3, \ldots\}$

$C^{(r)}{}_L(a^2) = \{ba^2, \ldots\}$

2) если $i \geq j$, то $C^{(r)}{}_L(a^i b a^j) = \{a^{i-j}\}$, число вхождений b равно 1; для разных i множества разные, число таких множеств счётно;

$C^{(r)}{}_L(b, aba, a^2ba^2, \ldots) = \{\varepsilon\}$

$C^{(r)}{}_L(ab, a^2ba, a^3ba^2, \ldots) = \{a\}$

$C^{(r)}{}_L(a^2b, a^3ba, \ldots) = \{a^2\}$

3) если $i < j$, то $C^{(r)}{}_L(a^iba^j) = \varnothing$, число вхождений b равно 1;

4) если $|y|_b > 1$, то $C^{(r)}{}_L(y) = \varnothing$, число вхождений b больше 1.

Рассмотрим **другой пример** $L = a^+b^*$. Для него

$C^{(r)}{}_L(\varepsilon) = \{a^+b^*\}$, различается с классом $C^{(r)}{}_L(a)$ словом b;

$C^{(r)}{}_L(a^i) = \{a^*b^*\}$, совпадает с классом $C^{(r)}{}_L(aa)$, различается с классом $C^{(r)}{}_L(\varepsilon)$ словом b;

$C^{(r)}{}_L(b\{a,b\}^* \cup a^+b^+a\{a, b\}^*) = \varnothing$, различается с классом $C^{(r)}{}_L(\varepsilon)$ словом a, с классом $C^{(r)}{}_L(a)$ словом a, с классом $C^{(r)}{}_L(ab)$ словом ε;

$C^{(r)}{}_L(a^+b^+) = \{b^*\}$, различается с классом $C^{(r)}{}_L(\varepsilon)$ словом a, с классом $C^{(r)}{}_L(b)$ словом ε.

Пусть дан язык $L \subseteq \Sigma^*$. Множество слов $W \subseteq \Sigma^*$ называется *базисом отношения* $\sim(L)$, если

а) любые два слова из W различимы относительно L,

б) любое слово из Σ^* эквивалентно некоторому слову из W.

Теорема 1. Пусть множество W попарно различимых относительно L слов обладает свойствами

1) $\varepsilon \in W$,

2) $\forall p \in W \quad \forall x \in \Sigma \quad \exists q \in W \quad px \sim q(L)$.

Тогда W – базис отношения $\sim(L)$.

Доказательство. Достаточно показать, что любое слово эквивалентно некоторому слову из W. Применим индукцию по длине слов. Для слов длины 0, т.е. ε, утверждение верно в силу 1) – любое слово $p \in W$ эквивалентно самому себе, т.е. слову $p\varepsilon$. Пусть оно верно для слов длины k. Рассмотрим произвольное слово s длины $k+1$. Представим его в виде $s = s`x$, $x \in \Sigma$. По предположению индукции $s` \sim q$ для некоторого $q \in W$. Тогда $s = s`x \sim qx$. Для слова qx ввиду 2) найдётся слово $r \in W$ такое, что $qx \sim r$. Поскольку отношение $\sim$ транзитивно, то $s \sim qx$, $qx \sim r \Rightarrow s \sim r \in W$, что и завершает доказательство. ♦

Опишем некоторый систематический способ построения классов эквивалентности. Метод нахождения базиса и одновременного построения таблицы функции перехода автомата заключается в следующем.

Строки таблицы соответствуют базисным словам, столбцы – буквам алфавита. Вначале имеется одно базисное слово ε и таблица имеет одну незаполненную строку. Опишем общий шаг заполнения таблицы. Находим первую сверху незаполненную строку и в ней первую слева пустую клетку. Пусть строка соответствует слову p, а клетка стоит в столбце, соответствующем букве x. Проверяем, есть ли среди слов, уже зачисленных в базис, слово, эквивалентное px. Если есть, вписываем его в указанную пустую клетку, а если нет, вписываем в неё px. Объявляем слово px базисным и добавляем соответствующую ему строку к таблице. Если на некотором шаге все строки таблицы оказались заполненными, то построение закончено. В том случае, когда построение бесконечно, – число классов бесконечно.

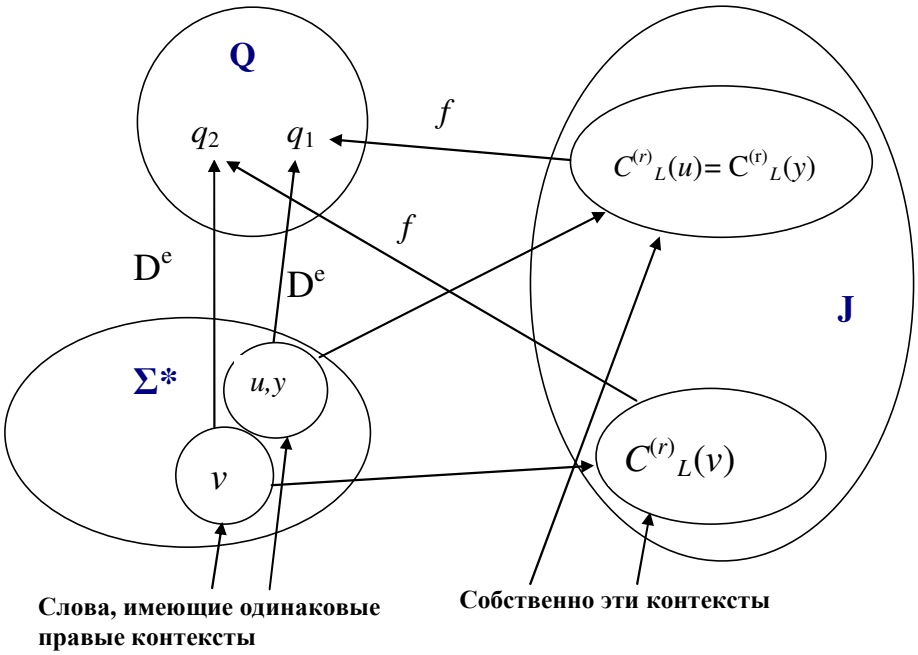
Для языка $L=\{a^{+}b^{*}\}$ получаем следующую таблицу:

	a	b
ε	a	b
a	a	ab
b	b	b
ab	b	ab

Лемма 2. Если язык L распознаётся полным детерминированным конечным автоматом $M = (Q, \Sigma, D, q_0, F)$, то $|\{C^{(r)}{}_L(y) \mid y \in \Sigma^*\}| \le |Q|$.

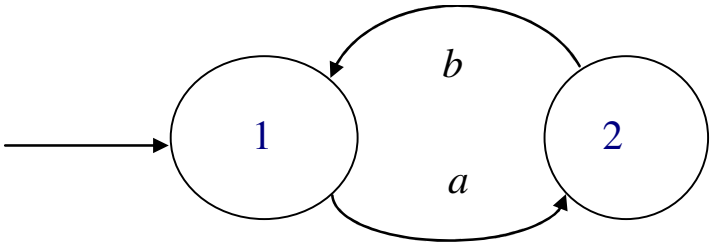
Доказательство. Введём обозначение $J \cong \{C^{(r)}{}_L(y) \mid y \in \Sigma^*\}$ – совокупность множеств правых контекстов языка L. Определим функцию f: $J \to Q$, положив $f(A)$ равным $D^e(q_0,y)$, где y — некоторое слово, для которого выполнено условие $C^{(r)}{}_L(y) = A$, т.е. это множество правых контекстов для $[y]$. Совокупность этих множеств и есть J. Заметим, что для любых слов u и v, если $C^{(r)}{}_L(u) \ne C^{(r)}{}_L(v)$, то $D^e(q_0,u) \ne D^e(q_0,v)$ (потому, что если бы было $D^e(q_0,u) = D^e(q_0, v)$, т.е. автомат по прочтении слов u и v оказывался в одном и том же состоянии, то и множество суффиксов z этих слов таких, что $uz \in L$ и $vz \in L$ совпадало бы). ♦

Следовательно, для разных элементов J образы функции f разные, т.е. функция f является инъективной. Но тогда $|J| \le |Q|$.



Пример 2. Пусть M = (Q, Σ, D, q_0, F), где $Q = \{1, 2\}$, $\Sigma = \{a, b\}$, $D = \{D(1, a) = 2, D(2, b) = 1\}$, $q_0 = 1$ и $F = \{1, 2\}$.

Тогда $L(M) = \{(ab)^n \mid n \ge 0\} \cup \{(ab)^n a \mid n \ge 0\}$.



Тогда $\{C^{(r)}{}_L(y) \mid y \in \Sigma^*\} = \{C^{(r)}{}_L(\varepsilon), C^{(r)}{}_L(a), C^{(r)}{}_L(b)\}$, $C^{(r)}{}_L(\varepsilon) = \{(ab)^n \mid n \ge 0\} \cup \{(ab)^n a \mid n \ge 0\}$, $C^{(r)}{}_L(a) = \{b(ab)^n \mid n \ge 0\} \cup \{(ba)^n \mid n \ge 0\}$ и $C^{(r)}{}_L(b) = \varnothing$.

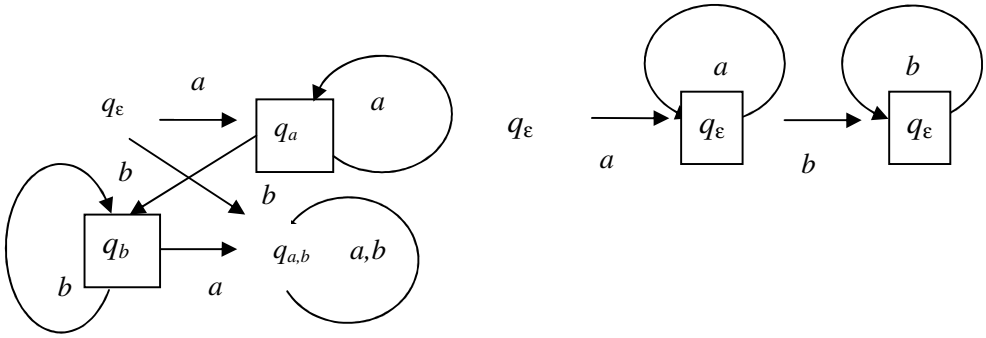
Лемма 3. Если $L \subseteq \Sigma^*$ и множество $\{C^{(r)}{}_L(y) \mid y \in \Sigma^*\}$ конечно, то язык L является регулярным.

Доказательство. Язык L распознаётся полным детерминированным конечным автоматом $M = (Q, \Sigma, D, q_0, F)$, где $Q = \{C^{(r)}{}_L(y) \mid y \in \Sigma^*\}$, $q_0 = C^{(r)}{}_L(\varepsilon)$, $F = \{C^{(r)}{}_L(y) \mid y \in L\}$, $\{D(C^{(r)}{}_L(y), a) = C^{(r)}{}_L(ya) \mid y \in \Sigma^*, a \in \Sigma\}$. ♦

Если базис вычислен и каждому базисному слову p и каждой букве $x \in \Sigma$ сопоставлено базисное слово q, где $q \sim px$, то тем самым определена функция перехода автомата $D([p], x) = [px] = [q]$.

Пример 3. Пусть $\Sigma = \{a, b\}$. Рассмотрим регулярный язык $L = a^+b^*$.

Обозначим $q_\varepsilon = C^{(r)}{}_L(\varepsilon)$, $q_a = C^{(r)}{}_L(a)$, $q_b = C^{(r)}{}_L(b) = \varnothing$, $q_{ab} = C^{(r)}{}_L(ab)$. Тогда $\{C^{(r)}{}_L(y) \mid y \in \Sigma^*\} = \{q_\varepsilon, q_a, q_b, q_{ab}\}$. Язык L распознаётся полным детерминированным конечным автоматом $\langle Q, \Sigma, D, q_0, F\rangle$, где $Q = \{q_\varepsilon, q_a, q_b, q_{ab}\}$, $q_0 = q_\varepsilon$, $F = \{q_a, q_{ab}\}$, $\{D(q_\varepsilon, a) = q_a, D(q_\varepsilon, b) = q_{ab}, D(q_a, a) = q_a, D(q_a, b) = q_b, D(q_b, a) = q_{ab}, D(q_b, b) = q_b, D(q_{ab}, a) = q_{ab}, D(q_{ab}, b) = q_{ab}\}$.



Теорема 2. Язык $L \subseteq \Sigma^*$ является регулярным тогда и только тогда, когда множество $\{C^{(r)}{}_L(y) \mid y \in \Sigma^*\}$ конечно.

Доказательство. Необходимость доказана в лемме 1, достаточность — в лемме 2. ♦

Замечание. В силу леммы 2 полный детерминированный конечный автомат, построенный в доказательстве леммы 3, является минимальным (по количеству состояний) среди всех полных детерминированных конечных автоматов, распознающих заданный язык, поскольку множество правых контекстов зависит только от языка и не может быть уменьшено, а автомат произвольный.