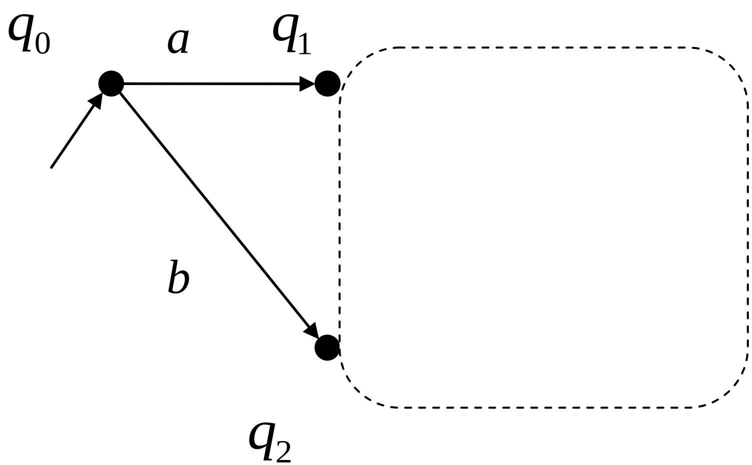
§6. Поиск регулярного выражения по автомату. Теорема Ардена. Решение системы уравнений с языками.

Задача: найти регулярное выражение для языка L, допускаемого ДКА (возможно неполным).

Замечание: в доказательстве теоремы Клини нет алгоритма поиска регулярного языка по автомату.

Пример. $\Sigma = \{a, b\}$



Обозначим S_0 – язык слов автомата, где q_0 – начальное состояние;
S_1 – язык слов автомата, где q_1 – начальное состояние;
S_2 – язык слов автомата, где q_2 – начальное состояние.

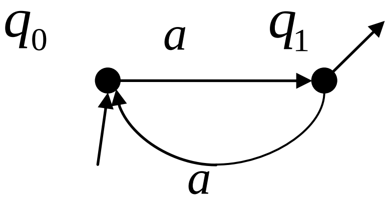
Тогда $S_0 = a \cdot S_1 + b \cdot S_2$.

Аналогично, S_1 и S_2 можно выразить через $\{S_i\}$, где
S_i – язык слов автомата, где q_i – начальное состояние.

Для заключительных состояний можно получить выражение, не содержащее S_i в правой части.

Проблемы возникают, когда для выражения S_i получаем в правой части тоже S_i .

Пример.



$$\begin{cases} S_0 = a \cdot S_1 \\ S_1 = a \cdot S_0 + \varepsilon \end{cases}$$

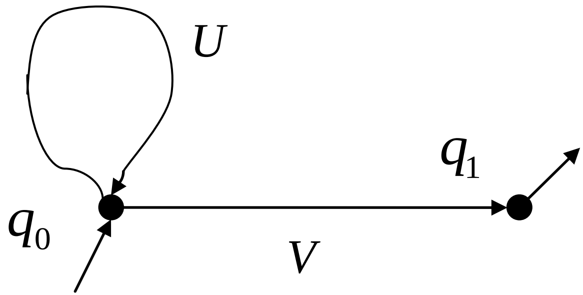
$$\Rightarrow \; S_0 = a \cdot (a \cdot S_0 + \varepsilon).$$

Воспользуемся догадкой: $S_0 = a \cdot \left(a^2\right)^* = \left(a^2\right)^* \cdot a.$

Теорема (Ардена).
Для любых регулярных языков U, X, V: $\varepsilon \notin U$, уравнение $X = U \cdot X + V$ имеет единственное решение $X = U^* \cdot V$.

Замечание: уравнению $X = X + V$ удовлетворяет любой язык X, для которого $V \subseteq X$.

Иллюстрация для запоминания теоремы Ардена:



Доказательство теоремы Ардена:

1) Подставим $X = U^* \cdot V$ в уравнение:

$U^* \cdot V = U \cdot U^* \cdot V + V$,

$U^* \cdot V = (U \cdot U^* + \varepsilon) \cdot V$ – верное равенство.

2) Пусть X' удовлетворяет уравнению $X = U \cdot X + V$.

Тогда $X' = U \cdot (U \cdot X' + V) + V = U^2 \cdot X' + (U + \varepsilon) \cdot V$.

$X' = U^2 \cdot (U \cdot X' + V) + V = U^3 \cdot X' + (U^2 + U + \varepsilon) \cdot V$.

$$\forall n \in \mathrm{N}$$

$$X' = U^{n+1} \cdot X' + (U^n + \ldots + U + \varepsilon) \cdot V \,.$$

$$\forall w \in U^* \cdot V \rightarrow \; w \in X' \,.$$

Осталось показать, что $X' \subseteq U^* \cdot V$.

(от противного) Предположим, $\exists w \in X' \colon\ w \notin U^* \cdot V$, w – кратчайшее.

$w \in UX',\ \ w \notin V$.

Тогда $w = u \cdot w'$, где $u \neq \varepsilon$, $u \in U$, $w' \in X'$.

$|w'| < |w| \ \Rightarrow w' \in U^* \cdot V$.

Тогда $w \in U^n \cdot V \subseteq U^* \cdot V$ – противоречие.

Следовательно, $X = U^* \cdot V$ – единственное решение. Теорема доказана.

Алгоритм поиска регулярного выражения:

1) Составим систему уравнений для S_i :

$$S_i = \sum_{a_j \in \Sigma} a_j \cdot S_m$$, если $q_i \notin Q_F$;

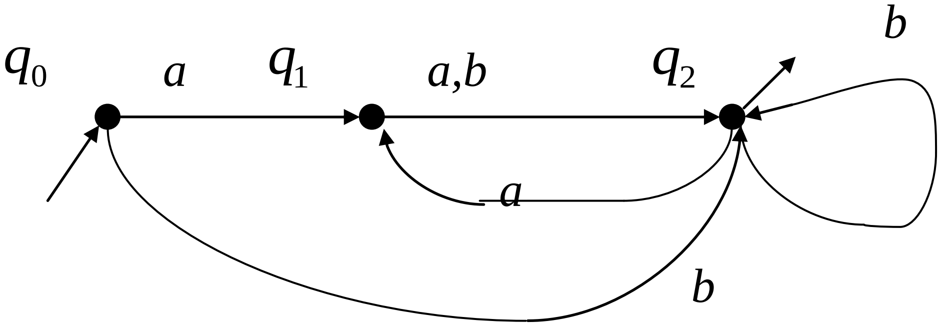
$$S_i = \varepsilon + \sum_{a_j \in \Sigma} a_j \cdot S_m$$, если $q_i \in Q_F$;

где $\varphi(q_i, a_j) = q_m$.

2) Удаляем последовательно S_i из системы.

3) S_0 – ответ.

Пример. Автомат для продажи кофе, где монета 5 соответствует символу *a*, монета 10 соответствует символу *b*.



$$\begin{cases} S_0 = a \cdot S_1 + b \cdot S_2 \\ S_1 = (a+b) \cdot S_2 \\ S_2 = \varepsilon + a \cdot S_1 + b \cdot S_2 \end{cases}$$

$$\begin{cases} S_0 = a \cdot S_1 + b \cdot S_2 \\ S_1 = (a+b) \cdot S_2 \\ S_2 = \varepsilon + a \cdot S_1 + b \cdot S_2 \end{cases}$$

Подставим S_1 из второго уравнения в первое и третье.

Получим $\begin{cases} S_0 = a \cdot (a+b) \cdot S_2 + b \cdot S_2 \\ S_2 = \varepsilon + a \cdot (a+b) \cdot S_2 + b \cdot S_2 \end{cases}$.

Упростим $\begin{cases} S_0 = (a^2 + ab + b) \cdot S_2 \\ S_2 = \varepsilon + (a^2 + ab + b) \cdot S_2 \end{cases}$

Выразим S_2 из второго уравнения: $S_2 = (a^2 + ab + b)^* \cdot \varepsilon$.

Это равносильно $S_2 = (a^2 + ab + b)^*$.

Подставим S_2 в первое уравнение, получим

$S_0 = (a^2 + ab + b) \cdot (a^2 + ab + b)^*$.

Или в более коротком виде: $S_0 = (a^2 + ab + b)^n$ – ответ.

В оригинальной статье Ардена использовалась альтернативная формулировка теоремы при альтернативных обозначениях.

Пусть T_i – язык слов автомата, где q_i – единственное заключительное состояние. Тогда уравнение с языками имеет вид:

$T_i = \sum_{a_j \in \Sigma} T_m \cdot a_j$, если q_i не является начальным;

$T_i = \varepsilon + \sum_{a_j \in \Sigma} T_m \cdot a_j$, если q_i – начальное состояние;

где $\varphi(q_m, a_j) = q_i$.

Альтернативная теорема Ардена.
Для любых регулярных языков U, X, V: $\varepsilon \notin U$, уравнение $X = V + X \cdot U$ имеет единственное решение

$$X = V \cdot U^{*}.$$

После решения системы уравнений нужно найти T_i для каждого состояния из Q_F. Ответом будет их сумма.