3. Крайни автомати. Регулярни изрази. Теорема на Клини.

Краен детерминиран автомат наричаме петорката

A = < **Q**, **X**, q_0, δ, **F** >, където

Q е крайно множество от вътрешни състояния,

X е крайна входна азбука,
$q_0 \in$ **Q** е начално състояние,

δ : **Q** x **X** → **Q** е частична **функция на преходите**,

F ⊆ **Q** е множеството от заключителни състояния.

Представяме крайните детерминирани автомати с краен ориентиран мултиграф с върхове елементите от **Q**. В графа има ребро от q_i до q_j, надписано с x ∈ **X,** ако δ (q_i, x) = q_j. Обикновено началното състояние е посочено със стрелка, а заключителните състояния се изобразяват с двойни кръгчета.

Дефинираме Δ : **Q** x **X*** → **Q** – **разширена функция на преходите**

по следния начин: Δ (q, ε) = q за всяко q ∈ **Q**,

Δ (q, αx) = δ (Δ (q, α), x), за всеки x ∈ **X**, q ∈ **Q** и α ∈ **X***, за които

δ (Δ (q, α), x) е дефинирана, недефинирана, ако δ (Δ (q, α), x) не е дефинирана.

Казваме, че автоматът **A** = < **Q**, **X**, q_0, δ, **F** > **разпознава думата**

α ∈ **X***, ако Δ (q_0, α) ∈ **F**. Езикът L_A = { α | α ∈ **X***, Δ (q_0, α) ∈ **F** }

наричаме **език, разпознаван от автомата A**.

Ако функцията на преходите δ : **Q** x **X** → **Q** не е тотална, то можем да разширим автомата **A** до **A′** следния начин:

A′ = < **Q** ∪ q*, **X**, q_0, **δ***, **F** >, където q* ∉ **Q**,

$\boldsymbol{\delta^*}$ (q, x) = $\boldsymbol{\delta}$ (q, x), ако $\boldsymbol{\delta}$ (q, x) е дефинирана и

$\boldsymbol{\delta^*}$ (q, x) = q*, ако $\boldsymbol{\delta}$ (q, x) не е дефинирана.

Сега, ако **A** разпознава думата α, то очевидно **A′** разпознава α.

Обратното също е вярно, тъй като $q^* \notin \mathbf{F} \Rightarrow L_A = L_{A'}$. И така без да изменяме езика на автомата можем, ако е необходимо, да додефинираме функцията на преходите до тотална функция.

Теорема: За всеки краен детерминиран автомат

A = < **Q**, **X**, q_0, δ, **F** > съществува автоматна граматика Γ, такава че

$L_Γ = L_A$.

Доказателство: Построяваме граматиката Γ = < **Q**, **X**, q_0, **P** >, където **P** = { $q_i \to xq_j$ | ако δ (q_i, x) = q_j } ∪ { $q_i \to x$, ако δ (q_i, x) ∈ **F** }.

Изчистваме Γ от аксиома в дясна част на правило и ако ε ∈ L_A

(q_0 ∈ **F**), добавяме правилото $q_0 \to ε$. Лесно се вижда, че получената граматика поражда езика L_A.

Краен недетерминиран автомат наричаме петорката

A = < **Q**, **X**, q_0, δ, **F** >, където

Q е крайно множество от вътрешни състояния,

X е крайна входна азбука,
q_0 ∈ **Q** е начално състояние,

δ : **Q** x **X** → $\mathbf{2^Q}$ е **функция на преходите**,

F ⊆ **Q** е множеството от заключителни състояния.

Дефинираме Δ : **Q** x **X*** → $\mathbf{2^Q}$ – **разширена функция на преходите**

по следния начин: Δ (q, ε) = { q } за всяко q ∈ **Q**,

Δ (q, x) = δ (q, x) за всяко x ∈ **X**, $Δ (q, αx) = \bigcup_{i=1}^{m} δ(q_{p_i}, x)$,

където $Δ (q, α) = \{ q_{p_1}, q_{p_2}, …, q_{p_m} \}$.

Казваме, че автоматът **A разпознава думата** α ∈ **X***, ако

$\Delta(q_0, \alpha) \cap \mathbf{F} \neq \varnothing$. Езикът $L_A = \{ \alpha \mid \alpha \in \mathbf{X^*}, \Delta(q_0, \alpha) \cap \mathbf{F} \neq \varnothing \}$

наричаме **език, разпознаван от автомата A**.

Теорема: За всяка автоматна граматика $\Gamma = < \mathbf{N}, \mathbf{T}, \mathbf{S}, \mathbf{P} >$ съществува краен недетерминиран автомат **A**, такъв че $L_A = L_\Gamma$.
Доказателство: Без ограничение на общността можем да считаме, че в Г няма преименуващи правила. Нека $E \notin \mathbf{N} \cup \mathbf{T}$. Конструираме автомата **A** по следния начин:

$\mathbf{A} = < \mathbf{N} \cup \{ E \}, \mathbf{T}, \mathbf{S}, \delta, \mathbf{F} >$, където

$\delta(A, x) = \{ B \mid A \rightarrow xB \in \mathbf{P} \} \cup \{ E \}$, ако $A \rightarrow x \in \mathbf{P}$ и
$\mathbf{F} = \{ E \}$, ако $\varepsilon \notin L_\Gamma$ и $\mathbf{F} = \{ \mathbf{S}, E \}$, ако $\varepsilon \in L_\Gamma$.
Тривиално се проверява, че $L_A = L_\Gamma$.

Теорема: За всеки краен недетерминиран автомат

$\mathbf{A} = < \mathbf{Q}, \mathbf{X}, q_0, \delta, \mathbf{F} >$ съществува краен детерминиран автомат $\mathbf{A'}$,

такъв че $L_A = L_{A'}$.

Доказателство: Построяваме крайният детерминиран автомат $\mathbf{A'}$

по следния начин: $\mathbf{A'} = < \mathbf{Q'}, \mathbf{X}, \{ q_0 \}, \delta', \mathbf{F'} >$, където

$\mathbf{Q'} \subseteq \mathbf{2^Q}$ е множеството от всички $M \subseteq \mathbf{Q}$, такива че съществува дума $\alpha \in \mathbf{X^*}$, такава че $\Delta(q_0, \alpha) = M$. Функцията $\delta' : \mathbf{Q'} \times \mathbf{X} \rightarrow \mathbf{Q'}$ е дефинирана по следния начин: $\delta'(\{q_{p_1}, q_{p_2}, \ldots, q_{p_m}\}, x) = \bigcup_{i=1}^{m} \delta(q_{p_i}, x)$

и $\mathbf{F'} = \{ M \mid M \in \mathbf{Q'}, M \cap \mathbf{F} \neq \varnothing \}$.

С индукция по дължината на α ще покажем, че

$\Delta(q_0, \alpha) = \Delta'(\{q_0\}, \alpha)$.

База: Нека $\alpha = \varepsilon$, т.е. $d(\alpha) = 0$. Имаме $\Delta(q_0, \varepsilon) = \{q_0\}$ и

$\Delta'(\{q_0\}, \varepsilon) = \{q_0\} \Rightarrow \Delta(q_0, \alpha) = \Delta'(\{q_0\}, \alpha)$.

Предположение: Нека твърдението е изпълнено за думата α.
Стъпка: Ще покажем, че $\Delta(q_0, \alpha x) = \Delta'(\{q_0\}, \alpha x)$.

Нека $\Delta'(\{q_0\}, \alpha) = \{q_{p_1}, q_{p_2}, \ldots, q_{p_m}\}$. Тогава по дефиницията на δ'

имаме $\Delta'(\{q_0\}, \alpha x) = \delta'(\{q_{p_1}, q_{p_2}, \ldots, q_{p_m}\}, x) = \bigcup_{i=1}^{m} \delta(q_{p_i}, x)$.

От индукционното предположение имаме

$\Delta(q_0, \alpha) = \Delta'(\{q_0\}, \alpha) = \{q_{p_1}, q_{p_2}, \ldots, q_{p_m}\} \Rightarrow$

$\Delta(q_0, \alpha x) = \bigcup_{i=1}^{m} \delta(q_{p_i}, x) = \Delta'(\{q_0\}, \alpha x)$.

Имаме $\Delta(q_0, \alpha) \cap \mathbf{F} \neq \varnothing \Leftrightarrow \Delta'(\{q_0\}, \alpha) \cap \mathbf{F} \neq \varnothing \Leftrightarrow \Delta'(\{q_0\}, \alpha) \in \mathbf{F'}$.

И така $L_A = L_{A'}$.

Като комбинираме трите теореми получаваме, че автоматните езици (тези които се разпознават от автоматни граматики), са точно тези езици, които се разпознават от крайни детерминирани (недетерминирани) автомати.

Оттук нататък считаме, че всички автомати са навсякъде дефинирани.

Крайните детерминирани автомати **A** и **A′** наричаме

еквивалентни, ако $L_A = L_{A'}$.

Крайният детерминиран автомат $\mathbf{A_0}$, който разпознава автоматния език L се нарича **минимален** за този език L, ако за всеки автомат **A**, който е еквивалентен на $\mathbf{A_0}$ имаме $|\mathbf{Q_0}| \leq |\mathbf{Q}|$, където $\mathbf{Q_0}$ и **Q** са множествата от състоянията съоствeтно на $\mathbf{A_0}$ и на **A**.

Нека X е произволна азбука. Релацията $R \subseteq X^* \times X^*$ наричаме

дясно-инвариантна, ако от $(\alpha, \beta) \in R \Rightarrow (\alpha\gamma, \beta\gamma) \in R$

за всяка дума $\gamma \in X^*$.

Нека $\mathbf{A} = < \mathbf{Q}, \mathbf{X}, q_0, \delta, \mathbf{F} >$ е краен детерминиран автомат.
Определяме релацията $R_A \subseteq \mathbf{X}^* \times \mathbf{X}^*$ по следния начин:

$R_A = \{ (\alpha, \beta) \mid \Delta (q_0, \alpha) = \Delta (q_0, \beta) \}$.

Лема: R_A е релация на еквивалентност и е дясно-инвариантна.
Доказателство: Тривиална проверка.

Нека X е произволна азбука и $L \subseteq X^*$ е произволен език.
Дефинираме релацията $R_L \subseteq X^* \times X^*$ по следния начин:
$R_L = \{ (\alpha, \beta) \mid$ за всяка дума $\gamma \in X^*$, $\alpha\gamma$ и $\beta\gamma$

едновременно са в L или не са в L $\}$.

Теорема: Релацията R_L е релация на еквивалентност и е дясно-инвариантна.

Доказателство: Тривиална проверка.

Нека $\mathbf{A} = < \mathbf{Q}, \mathbf{X}, q_0, \delta, \mathbf{F} >$ е краен детерминиран автомат.
Състоянието $q \in \mathbf{Q}$ наричаме **достижимо**, ако съществува дума

$\alpha \in \mathbf{X}^*$, такава че $\Delta (q_0, \alpha) = q$. Ако за всяка дума $\alpha \in \mathbf{X}^*$ имаме

$\Delta (q_0, \alpha) \neq q$, казваме че състоянието q е **недостижимо**.
Ясно е, че отстраняването на недостижимите състояния не

променя езика, разпознаван от автомата.

Така можем да считаме, че автоматът **A** има само достижими

състояния. Всеки клас на еквивалентност на R_A съдържа думите от $\mathbf{X}^*$, които довеждат автомата до дадено негово състояние. Ще отбележим, че тъй като автоматът е навсякъде дефиниран всяка дума от $\mathbf{X}^*$ довежда автомата до някое състояние. Така, ако автоматът **A** е навсякъде дефиниран и има само достижими състояния, индексът на релацията R_A е краен - точно $|\mathbf{Q}|$.

Теорема: Нека $L \subseteq \mathbf{X}^*$. Релацията $R_L \subseteq \mathbf{X}^* \times \mathbf{X}^*$ има краен индекс $\Leftrightarrow$

L е автоматен език.

Доказателство: Нека L е автоматен език. Тогава съществува краен

детерминиран автомат $\mathbf{A} = < \mathbf{Q}, \mathbf{X}, q_0, \delta, \mathbf{F} >$, такъв че $L_A = L$.
Лесно се вижда, че ако $(\alpha, \beta) \in R_A$, то $(\alpha, \beta) \in R_L$. При това положение, всеки клас на еквивалентност на R_L се състои от класове на R_A и тъй като R_A има краен индекс, то R_L също има краен индекс.

Нека индексът на R_L е краен. Нека $[\alpha]$ е класът на еквивалентност на релацията R_L с представител $\alpha \in \mathbf{X}^*$. Да означим с **Q** множеството от класовете на еквивалентност на R_L (то е крайно). Построяваме краен детерминиран автомат

$\mathbf{A} = < \mathbf{Q}, \mathbf{X}, [\varepsilon], \delta, \mathbf{F} >$. Лесно се вижда, че всеки клас на еквивалентност на релацията R_L съдържа думи само от L или думи не от L (от дефиницията на R_L при $\gamma = \varepsilon$).

Така можем да дефинираме $\mathbf{F} = \{ [\alpha] \mid \alpha \in L \}$.

Дефинираме $\delta([\alpha], x) = [\alpha x]$ за всяко $x \in \mathbf{X}$. Тази дефиниция е коректна, т.е. не зависи от представителя на класа на еквивалентност, тъй като релацията R_L е дясно-инвариантна. Лесно се вижда, че $L_A = L$. Така L е автоматен език.

Следствие: Автоматът $\mathbf{A} = < \mathbf{Q}, \mathbf{X}, [\varepsilon], \delta, \mathbf{F} >$ от теоремата е минимален за езика L.

Доказателство: В процеса на доказателството на теоремата показахме, че индексът на R_L е по-малък от индекса на R_A за всеки детерминиран автомат **A**, който разпознава L. От друга страна, построеният автомат има точно толкова състояния, колкото е индексът на R_L, така че той е минимален.

Нека **A** = < **Q**, **X**, q_0, δ, **F** > е краен детерминиран автомат и L_A = L. Считаме, че автоматът е навсякъде дефиниран и има само достижими състояния.

Състоянията q_i, $q_j \in$ **Q** наричаме **еквивалентни**, ако съответните им класове на еквивалентност на релацията R_A попадат в един и същ клас на еквивалентност на релацията R_L.

От тази дефиниция и от теоремата лесно се вижда, че минимизацията на автомата **A** се свежда до намиране на всички подмножества на **Q** от еквивалентни състояния.

Лема: Ако $q_1, q_2 \in$ **Q**, $q_1 \notin$ **F** и $q_2 \in$ **F**, то q_1 и q_2 не са еквивалентни.

Лема (тест на едната буква): Нека $q_1, q_2 \in$ **Q**. Ако съществува

$x \in$ **X**, така че δ (q_1, x) не е еквивалентно на δ (q_2, x), то q_1 и q_2 не са еквивалентни.

Лема: Нека $\mathcal{Q} = \{ \mathcal{Q}_1, \mathcal{Q}_2, \ldots, \mathcal{Q}_s\}$ е разбиване на **Q**, такова че

$\mathcal{Q}_j \subseteq$ **F** или $\mathcal{Q}_j \subseteq$ **Q****F**, j = 1, 2, …, s. Нека за всяко $\mathcal{Q}_j$, за всяка буква $x \in$ **X** и за всеки $q_1, q_2 \in \mathcal{Q}_j$ имаме: δ (q_1, x) $\in \mathcal{Q}_k$,

δ (q_2, x) $\in \mathcal{Q}_k$ за някое k $\in$ { 1, 2, …, s}. Тогава всяко $\mathcal{Q}_j$ се състои само от еквивалентни състояния.

Трите леми водят до следния алгоритъм:

1. Образуваме разбиването $S^0 = \{ Q_1^0, Q_2^0\}$, където Q_1^0 = **F** и Q_2^0 = **Q****F**, i = 0.
2. Нека сме построили разбиването $S^i = \{ Q_1^i, Q_2^i, \ldots, Q_{l_i}^i \}$.

   Всяко Q_j^i разбиваме на $\{ Q_{j_1}^{i+1}, Q_{j_2}^{i+1}, \ldots, Q_{j_m}^{i+1} \}$, такива че

   елементите на всяко от тях не се държат като нееквивалентни с теста на едната буква и обединяваме получените разбивания в S^{i+1}.

3. Ако $S^{i+1} = S^i$ – край, иначе i++, премини към 2.

Нека $S^r = S^{r+1} = \dots$ е последното разбиване.

Ясно е, че от последната лема подмножествата в S^r се състоят само

от еквивалентни състояния. Освен това лесно се вижда, че не е възможно еквивалентни състояния да са в различни подмножества на разбиването S^r.

Сега вече лесно можем да построим минималният автомат

$\mathbf{A_0} = < \mathbf{Q_0}, \mathbf{X}, t_0, \delta_\mathbf{0}, \mathbf{F_0} >$, $\mathbf{Q_0} = S^r$. Избираме $t_0 = Q^r_p$, така че $q_0 \in Q^r_p$,

тъй като началното състояние на автомата е класът на еквивалентност на R_L, който съдържа ε, а ε е в класът на еквивалентност на R_A, съответен на q_0. Множеството от заключителните състояния $\mathbf{F_0}$ определяме по следния начин:

$\mathbf{F_0} = \{ Q^r_p \mid Q^r_p \subseteq \mathbf{F} \}$, тъй като заключителните състояния са тези класове на еквивалентност на R_L, които съдържат само думи от L и тогава тези класове са образувани точно от класовете на R_A, съответни на заключителни състояния.

Функцията на преходите $\boldsymbol{\delta_0}$ дефинираме по следния начин:

$\boldsymbol{\delta_0}(Q^r_p, x) = Q^r_j$, ако за всяко $q \in Q^r_p$ имаме $\boldsymbol{\delta}(q, x) \in Q^r_j$.

Дефиницията е коректна, тъй като във всички елементи на S^r, състоянията реагират еднакво на теста на едната буква.

Нека X е азбука. Въвеждаме операции в множеството от всички езици над азбуката X:

1. Нека $L_1 \subseteq X^*$, $L_2 \subseteq X^*$. Дефинираме **сума** на езиците L_1 и L_2: $L_1 + L_2 = L_1 \cup L_2$.

2. Нека $L_1 \subseteq X^*$, $L_2 \subseteq X^*$. Дефинираме **произведение** на езиците L_1 и L_2: $L_1.L_2 = \{ \alpha\beta \mid \alpha \in L_1, \beta \in L_2 \}$.

   Дефинираме $L^0 = \{\varepsilon\}$, $L^1 = L$, $L^2 = LL$, …, $L^{n+1} = L^nL$.

3. Нека $L \subseteq X^*$. Дефинираме **итерация** на езика L: $L^* = \bigcup_{n=0}^{\infty} L^n$.

Нека $X = \{ x_1, x_2, \ldots, x_n\}$ е крайна азбука.

Разширяваме X до $\overline{X} = X \cup \{ \varepsilon, \varnothing, *, +, ., (,) \}$.

Без ограничение на общността можем да считаме, че добавените символи не се срещат в X. Дефинираме индуктивно понятията **регулярен израз** и **език, съпоставен на този регулярен израз**:

(регулярният израз е дума над $\overline{X}$, съпоставените езици са над X)

База: Думите ε, $\varnothing$, x_i, $i = 1, 2, \ldots, n$ над $\overline{X}$ са регулярни изрази и съответните им регулярни езици над X са

$\{ \varepsilon \}$, $\varnothing$, $\{ x_i \}$, $i = 1, 2, \ldots, n$.

Предположение: Нека $\alpha, \beta \in \overline{X}^*$ са регулярни изрази и съответните им регулярни езици над X са L_α и L_β.

Стъпка: Тогава $(\alpha)+(\beta)$, $(\alpha).(\beta)$, $(\alpha)^* \in X^*$ са регулярни изрази и

съответните им регулярни езици над X са $L_\alpha + L_\beta$, $L_\alpha.L_\beta$, L_α^*.

Въвеждаме приоритет на операциите в следния ред:

итерация, произведение и сума. По този начин можем съществено

да ограничим употребата на скоби в регулярните изрази.

Теорема: Ако L_1 и L_2 са автоматни езици, то $L_1 + L_2$ е автоматен език.

Доказателство:

Нека езикът L_1 се поражда от граматиката $\Gamma_1 = < \mathbf{N_1}, \mathbf{T_1}, \mathbf{S_1}, \mathbf{P_1} >$.
Нека езикът L_2 се поражда от граматиката $\Gamma_2 = < \mathbf{N_2}, \mathbf{T_2}, \mathbf{S_2}, \mathbf{P_2} >$.
Без ограничение на общността можем да смятаме, че

$\mathbf{N_1} \cap \mathbf{N_2} = \varnothing$, $\mathbf{N_1} \cap \mathbf{T_2} = \varnothing$, $\mathbf{T_1} \cap \mathbf{N_2} = \varnothing$. Ако това не е изпълнено можем да преименуваме съответните нетерминали, което няма да измени езиците L_1 и L_2. Нека $\mathbf{S} \notin \mathbf{N_1} \cup \mathbf{N_2} \cup \mathbf{T_1} \cup \mathbf{T_2}$.

Построяваме граматиката $\Gamma = < \mathbf{N_1} \cup \mathbf{N_2} \cup \{ \mathbf{S} \}, \mathbf{T_1} \cup \mathbf{T_2}, \mathbf{S}, \mathbf{P} >$,

$\mathbf{P} = (\mathbf{P_1} \cup \mathbf{P_2}) \setminus \{ \mathbf{S_1} \to \varepsilon, \mathbf{S_2} \to \varepsilon \} \cup \{ \mathbf{S} \to \mathbf{S_1}, \mathbf{S} \to \mathbf{S_2} \} \cup \{ \mathbf{S} \to \varepsilon$, ако имаме $\mathbf{S_1} \to \varepsilon$ или $\mathbf{S_2} \to \varepsilon \}$. Получената граматика е автоматна и

веднага се вижда, че $L_\Gamma = L_{\Gamma_1} + L_{\Gamma_2} = L_1 + L_2$.

Теорема: Всеки краен език е автоматен.

Доказателство: $\{ \varepsilon \}$ е автоматен език, тъй като се поражда от

автоматната граматика $\Gamma = < \{ \mathbf{S} \}, \{ a \}, \mathbf{S}, \{ \mathbf{S} \to \varepsilon \} >$.

Граматиката $\Gamma = < \{ \mathbf{S} \}, \{ a \}, \mathbf{S}, \{ \mathbf{S} \to a\mathbf{S} \} >$ е автоматна и поражда

езика $\varnothing$. Нека $L = \{ \alpha \}$, $\alpha \neq \varepsilon$, $\alpha = a_1 a_2 \ldots a_k$.

Тогава граматиката $\Gamma = < \{ \mathbf{S}, A_1, A_2, \ldots, A_{k-1} \}, \{ a_1, a_2, \ldots, a_k \}, \mathbf{S},$

$\{ \mathbf{S} \to a_1 A_1, A_1 \to a_2 A_2, \ldots, A_{k-2} \to a_{k-1} A_{k-1}, A_{k-1} \to a_k \} >$ е автоматна и очевидно $L_\Gamma = L = \{ \alpha \}$.

Нека $L = \{ \alpha_1, \ldots, \alpha_m \}$ е произволен краен език, $m \geq 1$.

Тогава $L = \{ \alpha_1 \} + \{ \alpha_2 \} + \ldots + \{ \alpha_m \}$ и от предната теорема L е автоматен език, тъй като е крайна сума на автоматни езици.

Теорема: Ако L_1 и L_2 са автоматни езици, то $L_1.L_2$ е автоматен език.

Доказателство:

Нека езикът L_1 се поражда от граматиката $\Gamma_1 = < \mathbf{N_1}, \mathbf{T_1}, \mathbf{S_1}, \mathbf{P_1} >$.
Нека езикът L_2 се поражда от граматиката $\Gamma_2 = < \mathbf{N_2}, \mathbf{T_2}, \mathbf{S_2}, \mathbf{P_2} >$.
Отново без ограничение на общността можем да смятаме, че

$\mathbf{N_1} \cap \mathbf{N_2} = \varnothing$, $\mathbf{N_1} \cap \mathbf{T_2} = \varnothing$, $\mathbf{T_1} \cap \mathbf{N_2} = \varnothing$.
Нека $\varepsilon \notin L_1$, $\varepsilon \notin L_2$.

Построяваме граматиката $\Gamma = < \mathbf{N_1} \cup \mathbf{N_2}, \mathbf{T_1} \cup \mathbf{T_2}, \mathbf{S_1}, \mathbf{P} >$,

$\mathbf{P} = (\mathbf{P_1} \cup \mathbf{P_2}) \setminus \{ A \to a \mid A \to a \in \mathbf{P_1} \} \cup \{ A \to a\mathbf{S_2} \mid$ за всяко

$A \to a \in \mathbf{P_1} \}$. Очевидно Γ е автоматна граматика и веднага се проверява, че $L_\Gamma = L_1.L_2$.

Нека $\varepsilon \in L_1$, $\varepsilon \notin L_2$ (случаят $\varepsilon \notin L_1$, $\varepsilon \in L_2$ се разглежда аналогично).
Тогава $L_1 = L_1' + \{ \varepsilon \}$, $\varepsilon \notin L_1'$, и L_1' също е автоматен език.

Сега $L_1.L_2 = (L_1' + \{ \varepsilon \}).L_2 = L_1'.L_2 + L_2$. От токущо доказаното $L_1'.L_2$ е автоматен език, L_2 е автоматен език и от теоремата по-горе $L_1.L_2$ е

автоматен език.

Нека $\varepsilon \in L_1$, $\varepsilon \in L_2$. Тогава $L_1 = L_1' + \{ \varepsilon \}$ и $\varepsilon \notin L_1'$, $L_2 = L_2' + \{ \varepsilon \}$ и

$\varepsilon \notin L_2'$, L_1' и L_2' също са автоматни езици.

Имаме $L_1.L_2 = (L_1' + \{ \varepsilon \}).(L_2' + \{ \varepsilon \}) = L_1'.L_2' + L_1' + L_2' + \{ \varepsilon \}$.

От доказаното по-горе $L_1'.L_2'$ е автоматен език, също L_1' и L_2' са автоматни езици и $\{ \varepsilon \}$ е автоматен език, защото е краен ➔
$L_1.L_2$ е автоматен език.

Теорема: Ако L е автоматен език, то L^* е автоматен език.
Доказателство:

Нека езикът L се поражда от граматиката $\Gamma = < \mathbf{N}, \mathbf{T}, \mathbf{S}, \mathbf{P} >$.
Без ограничение на общността можем да смятаме, че Г е без аксиома в дясна част на правило.

Построяваме граматика $\Gamma' = < \mathbf{N}, \mathbf{T}, \mathbf{S}, \mathbf{P'} >$, където

$\mathbf{P'} = \mathbf{P} \setminus \{ \mathbf{S} \rightarrow \varepsilon \} \cup \{ A \rightarrow a\mathbf{S} \mid$ за всяко късо правило $A \rightarrow a \in \mathbf{P} \}$.

Очевидно Γ′ е автоматна граматика, вижда се, че $L_{\Gamma'} = L^*\backslash\{\varepsilon\}$.

И така $L^*\backslash\{\varepsilon\}$ е автоматен език ➔ $L^* = L^*\backslash\{\varepsilon\} \cup \{\varepsilon\}$ е автоматен език.

Теорема (Клини): Множествата на регулярните езици и автоматните езици съвпадат.

Доказателство: С индукция по построението ще покажем,

че всеки регулярен език е автоматен.

База: Езиците $\{\varepsilon\}$, $\varnothing$, $\{x_i\}$ са автоматни езици, защото са крайни.
Предположение: Нека L_α и L_β са регулярни езици, съответни на регулярните изрази α и β и да допуснем, че L_α и L_β са автоматни езици.

Стъпка: Тогава $L_\alpha + L_\beta$, $L_\alpha.L_\beta$, L_α^* са автоматни езици

(теоремите по-горе).

Нека L е автоматен език. Ще докажем, че L е регулярен език.
Съществува краен детерминиран автомат

A = < **Q**, **X**, q_0, δ, **F** >, такъв че $L_A = L$. Да означим

Q = $\{q_0, q_1, \ldots, q_n\}$, **F** = $\{q_{p_1}, q_{p_2}, \ldots, q_{p_r}\}$ и да си мислим, че автоматът

A е представен с крайния ориентиран мултиграф G.

Означаваме с R^k_{ij} множеството от маршрутите в G от q_i до q_j, които

не използват като вътрешни върховете $q_k, q_{k+1}, \ldots, q_n$. Ясно е, че всеки маршрут от q_i до q_j определя дума $\alpha \in$ **X***,

такава че $\Delta(q_i, \alpha) = q_j$. Така можем да отъждествим всеки маршрут

със съответната му дума от входната азбука и да считаме, че R^k_{ij} е

език над **X**, т.е. $R^k_{ij} \subseteq$ **X***.

Лема: За всеки $q_i, q_j \in$ **Q** е в сила: $R^{k+1} = R^k + R^k.(R^k)^*.R^k$,

R Доказателство: Маршрутите от

k = 0, 1, …, n.

ij
ij
ik
kk
kj

k+1 ij

ра
зб
ив
ам
е
на
дв
е

подмножества – такива които не използват q_k като вътрешен връх и такива които използват q_k като вътрешен връх. Очевидно маршрутите от R_{ij}^{k+1}, които не използват q_k като вътрешен връх са точно маршрутите от R_{ij}^{k}.

Да разбием произволен маршрут от R_{ij}^{k+1}, който използва q_k като вътрешен връх на части по следния начин: q_i -- q_k -- q_k …q_k -- q_j.

В междинните части не се среща q_k, така че маршрутът започва с маршрут от q_i до q_k, който не използва q_k като вътрешен връх, т.е. маршрут от R_{ik}^{k}, продължава с произволен брой цикли от q_k до q_k, които минават точно веднъж през q_k, т.е. маршрут от $(R_{kk}^{k})^*$ и завършва с маршут от q_k до q_j, който не използва q_k като вътрешен връх, т.е. маршрут от R_{kj}^{k}. Така маршрутите от R_{ij}^{k+1}, които използват q_k като вътрешен връх са точно $R_{ik}^{k}.(R_{kk}^{k})^*.R_{kj}^{k}$ и окончателно $R_{ij}^{k+1} = R_{ij}^{k} + R_{ik}^{k}.(R_{kk}^{k})^*.R_{kj}^{k}$.

Лема: Езикът R_{ij}^{k} е регулярен език за всеки i, j $\in \{0, 1, \ldots, n\}$, $k \in \{0, 1, \ldots, n+1\}$.

Доказателство: Провеждаме индукция по k.

База: k = 0. Ако i = j, то R_{ii}^{0} са всички маршрути от q_i до q_i, които не минават през никой друг връх. Ако в графа G няма примки в q_i, тогава $R_{ii}^{0} = \{\varepsilon\}$ и R^{0} е регулярен език. Ако в графа G има примки в q_i и на тях съответстват букви $x_{i_1}, x_{i_2}, \ldots, x_{i_r}$, то

$R_{ii}^{0} = \{x_{i_1}, x_{i_2}, \ldots, x_{i_r}, \varepsilon_{ii}\}$ и R_{ii}^{0} е регулярен език, тъй като съответства на регулярния израз $x_{i_1} + x_{i_2} + \ldots + x_{i_r} + \varepsilon$.

Ако i $\neq$ j, то R_{ij}^{0} са ребрата от q_i до q_j. Ако в графа G няма ребро от q_i до q_j, то $R_{ij}^{0} = \varnothing$ и R_{ij}^{0} е регулярен език. Ако в графа G има ребра от q_i до q_j и на тях съответстват букви $x_{i_1}, x_{i_2}, \ldots, x_{i_r}$, то

$R_{ij}^{0} = \{x_{i_1}, x_{i_2}, \ldots, x_{i_r}\}$ и R_{ij}^{0} е регулярен език, тъй като съответства на регулярния израз $x_{i_1} + x_{i_2} + \ldots + x_{i_r}$.

Предположение: Нека за някое k езикът R_{ij}^{k} е регулярен език и нека α_{ij}^{k} е съответният регулярен израз.

Стъпка: Тогава езикът R_{ij}^{k+1} също е регулярен, тъй като съгласно горната лема този език съответства на регулярния израз

$\alpha_{ij}^{k} + \alpha_{ik}^{k}.(\alpha_{kk}^{k})^{*}.\alpha_{kj}^{k}.$

Продължение на доказателството на теоремата на Клини:

Имаме, че R_{ij}^{n+1} са всички маршрути от q_i до q_j (без ограничение) и

от горната лема R_{ij}^{n+1} е регулярен език. От дефиницията за езика L_A имаме, че $L_A = R_{0p_1}^{n+1} + R_{0p_2}^{n+1} + \ldots + R_{0p_r}^{n+1}$, тъй като думите, които се разпознават от автомата го довеждат от началното състояние

q_0 до някое от крайните $q_{p_1}, q_{p_2}, \ldots, q_{p_r}$. И така L_A е крайна сума на

регулярни езици ➔ $L = L_A$ е регулярен език. Така всеки автоматен език е регулярен.