Множина всіх слів в алфавіті Σ позначається через Σ^*. Множина всіх непустих слів в алфавіті Σ позначається через Σ^+.

Мова над алфавітом Σ – це підмножина Σ^*, тобто це множина ланцюжків скінченної довжини у заданому скінченному алфавіті. Наприклад, множина $L = \{a, abb\}$ є мовою над алфавітом $\Sigma = \{a, b\}$

Оскільки кожна мова є множиною, можна розглядати теоретико-множинні операції **об'єднання, перетину, різниці мов** (позначення $L_1 \cup L_2$, $L_1 \cap L_2$, $L_1 - L_2$), заданих над одним і тим самим алфавітом.

Нехай $L \subseteq \Sigma^*$. Тоді мова $\Sigma^* - L$ є **доповненням** мови L відносно алфавіту Σ. Коли з контексту зрозуміло, про який алфавіт йдеться, говорять просто, що мова $\Sigma^* - L$ являється доповненням мови L.

Теорема Кліні

Мова L регулярна тоді і тільки тоді, коли вона є автоматною.

Доводять теорему:

– визначенням алгоритмів побудови за регулярним виразом відповідного атомата і навпаки;

– за автоматом регулярного виразу.

Доповнення автоматної мови є автоматною мовою.

з теореми про доповнення автоматних мов та теореми Кліні:

операції об'єднання та доповнення регулярних мов не виводять за межі класу регулярних мов.

Для зручності опису формул прийнято таке позначення для операції доповнення $\overline{L}$ – доповнення до множини L.

Оскільки кожна автоматна мова є регулярною мовою (з теореми Кліні), а доповнення до автоматної мови є автоматною мовою, то доповнення до регулярної мови є регулярною мовою. Отже:

- *об'єднання регулярних мов є регулярною мовою;*
- *доповнення регулярних мов є регулярною мовою,*

тобто операції об'єднання та доповнення не виводять за межі класу регулярних мов.

Для множин справедливі такі співвідношення:

1. Перетин множин можна подати як

$$L_1 \cap L_2 = \overline{\overline{L_1} \cup \overline{L_2}}.$$

2. Різниця множин представляється як

$$L_1 \setminus L_2 = L_1 \cap \overline{L_2}.$$

Співвідношення 1 і 2 представляють перетин та різницю регулярних мов через операції доповнення та об'єднання. Оскільки перетин та різницю регулярних мов можна виразити через операції об'єднання та доповнення, то вони теж не виводять за межі класу регулярних мов.

Отже,

- *перетин регулярних мов є регулярною мовою;*
- *різниця регулярних мов є регулярною мовою.*

Тому

клас регулярних мов замкнений відносно операцій перетину та різниці.

Конкатенацією множин A та B (позначають $A \cdot B$ або AB) називають множину всіх ланцюжків вигляду $\alpha\beta$, де $\alpha \in A$ і $\beta \in B$.

Ітерацією множини A (позначають A^*) називають множину, визначену індуктивно так:

$A^0 = \varepsilon$;

$A^n = AA^{n-1}, n \geq 1$;

$A^* = \bigcup_{n=0}^{\infty} A^n$, де ε – порожній ланцюжок.

Приклад

Нехай $A = \{aa, bb\}$, $B = \{\varepsilon, a\}$.

Тоді $AB = \{aa, bb, aaa, bba\}$ і

$A^* = \{\varepsilon, aa, bb, aaaa, aabb, bbaa, bbbb, aaaaaa, \ldots\}$.

Регулярний вираз над алфавітом Σ визначається рекурсивно так:

0 є регулярним виразом;

1 є регулярним виразом;

a є регулярним виразом, якщо $a \in \Sigma$;

$(e+f)$,

$(e \cdot f)$

e^* є регулярними виразами, якщо e і f – регулярні вирази.

Для зменшення кількості дужок у виразах введено такий пріоритет раніше введених операцій. Визначено, що операція * має вищий пріоритет, ніж множення, а множення має вищий пріоритет, ніж додавання. Замість $e \cdot f$ часто пишуть просто ef.

Приклад

Нехай $\Sigma = \{a, b\}$. Тоді $((a \cdot b)^* \cdot (1+a))$ є регулярним виразом над алфавітом Σ.

Кожний регулярний вираз e над алфавітом Σ **задає мову** над алфавітом Σ (позначається $L(e)$ де e – регулярний вираз), яка визначається рекурсивно так, якщо e, f – регулярні вирази:

$$L(0) = \varnothing;$$

$$L(1) = \varepsilon;$$

$$L(a) = \{a\}, \text{ якщо } a \in \Sigma;$$

$$L(e + f) = L(e) \cup L(f);$$

$$L(e \cdot f) = L(e) \cdot L(f);$$

$$L(e^*) = L(e)^*.$$

Замість $L(e)$ часто пишуть просто e.

Приклад

Нехай $\Sigma = \{a, b\}$. Згідно з визначенням

$$L((a \cdot b)^* \cdot (1 + a)) = \{(ab)^n \mid n \geq 0\} \cup \{(ab)^n a \mid n \geq 0\}.$$

Для будь-яких регулярних виразів e, f, g виконуються такі тотожності:

$$e+f=f+e;$$

$$e+0=e;$$

$$e\cdot 1=1\cdot e=e;$$

$$e\cdot 0=0\cdot;$$

$$e+e=e$$

$$e+e^{*}=e^{*}$$

$$(e^{*})^{*}=e^{*}$$

$$(e+f)+g=e+(f+g);$$

$$(e\cdot f)\cdot g=e\cdot(f\cdot g);$$

$$e\cdot(f+g)=e\cdot f+e\cdot g;$$

$$(f+g)\cdot e=f\cdot e+g\cdot e;$$

$$(1+e+ee+\ldots+e^{n-1})(e^{n})^{*}=e^{*},\ \forall n\geq 1$$

$$(e^{*}+f)^{*}e^{*}=(e+f)^{*}$$

$$1+e(fe)^{*}f=(ef)^{*}.$$

Формування регулярного виразу для заданого скінченного автомату.

Алгоритм базується на побудові узагальненого скінченного автомата на основі заданого недетермінованого.

Узагальненим скінченним автоматом називається аналог скінченного автомата з одним вхідним та одним вихідним станом, де переходи помічені не словами, а регулярними виразами. *Мітка шляху* такого автомата – добуток регулярних виразів на переходах цього шляху.

Якщо є декілька переходів зі спільним початком і спільним кінцем (такі переходи називаються *паралельними*), замінимо їх на один перехід, використовуючи операцію “+”.