

Będziemy rozważać skończony, skierowany graf G . Wprowadźmy pewne pojęcia i oznaczenia:

- N - rozmiar (ilość symboli alfabetycznych)
- dla N i wyrażeń regularnych $R1$ i $R2$ mamy następujące zależności indukcyjne

	symbol z alfabetu (nie λ)	$R1 \cup R2$	$R1R2$	R^*
N	1	$N(R1) + N(R2)$	$N(R1) + N(R2)$	$N(R)$

- wiązka - zbiór ścieżek (t.j. połączonych sekwencji krawędzi) w grafie G , które mają wspólny wierzchołek startowy i wspólny wierzchołek końcowy
- pętla - ścieżka, której wierzchołek startowy = wierzchołkowi końcowemu
- punkt bazowy - jeśli wiązka składa się całkowicie z pętli, to jego wierzchołek początkowy (a zarazem końcowy) nazywamy punktem bazowym
- λ - pętla o długości zero na punkcie bazowym n
- Jeśli P i Q są wiązkami dla których wierzchołek końcowy dla wszystkich ścieżek w P = wierzchołkowi startowemu wszystkich ścieżek w Q , to definiujemy PQ , konkatenację P i Q , jako zbiór wszystkich ścieżek otrzymany jako poprzedzenie ścieżki z Q ścieżką z P
- Jeśli P jest wiązką z punktem bazowym n , definiujemy
 $P^* = \lambda_n$ suma P suma $PP \dots$
- Wiazkę będziemy nazywać regularną, jeśli będzie zbudowana z wiązek zawierających pojedyncze krawędzie przy wykorzystaniu sumy, konkatenacji i operatora $*$
- Istnieje jednoznaczność pomiędzy wiązkami, a wyrażeniami regularnymi. Wyrażenia reprezentowane przez wiązki stanowią podklasę wszystkich wyrażeń regularnych i mają własność, że każde podwyrażenie również reprezentuje wiązkę
- Wiazka S pokrywa ścieżkę p jeśli jest ścieżka q w S dla której p jest podścieżką.
- Jeśli istnieje n takie, że S pokrywa p^n , ale nie p^{n+1} wtedy to n oznaczamy jako $I_p(S)$ i nazywamy indeksem p w S oraz mówimy, że S jest p -skończona. W przeciwnym przypadku mówimy, że S jest p -nieskończona. Biorąc wcześniej zdefiniowane N : (dla wszystkich ścieżek p i wiązek R i S .)
 $I_p(\text{symbol z alfabetu}) = 0$ lub 1
 $I_p(R \cup S) = \max(I_p(R), I_p(S)) \rightarrow R \cup S$ jest określona tylko dla wiązek o wspólnym pocz. i końcu.
 $I_p(RS) \leq I_p(R) + I_p(S) + 1$

$$I_p(R^*) = \max I_p(R^k): k \geq 0$$

Porównując indukcyjne nierówności dla $1 + I_p$ i N mamy:

$$I_p(E) \leq 2N(E) \text{ jeśli } E \text{ jest p-skończona}$$

Jeśli E jest p-nieskończona, zbiór podwyrażeń E , które są p-nieskończone, jest niepusty i ma element minimalny ze względu na relację zawierania. Jak wynika z zależności indukcyjnych indeksu, każde minimalne p-nieskończone podwyrażenie jest postaci F^* , gdzie F jest p-skończona.

Dowód:

Pokażemy, że $2^{n-1} \leq N$ (dla pełnego grafu o n wierzchołkach)

Chcemy pokazać, że istnieje pętla p z punktem bazowym 1 w pełnym grafie z wierzchołkami 1, 2, ..., n , dla którego $N(\text{jakiegokolwiek wiązki pokrywającej } p) \geq 2^{n-1}$.

Będziemy przeprowadzać indukcję po n . Twierdzenie dla $n=1$ jest oczywiste, bierzemy jako pętlę wierzchołek p .

Założmy, że mamy pętlę p z punktem bazowym 1 w grafie pełnym z $n-1$ wierzchołkami dla którego $N(\text{jakiegokolwiek wiązki pokrywającej } p) \geq 2^{n-2}$.

W grafie pełnym o n wierzchołkach, zdefiniujemy p_k jako pętlę otrzymaną z p przez cykliczną permutację wierzchołków, która zastępuje wierzchołek 1 wierzchołkiem k . p_k ma punkt bazowy k , i omija wierzchołek $k-1$ (gdy $k=1$, wtedy omijamy wierzchołek n -ty). Rozważmy pętlę

$$q = p_1^m a_{12} p_2^m a_{23} \dots p_n^m a_{n1}$$

gdzie $m=2^n$ i a_{ij} jest krawędzią z wierzchołka i do j . Niech wyrażenie E reprezentuje wiązkę pokrywającą q . Dla każdego k , indeks p_k w $E \geq 2^n$ więc albo E jest p_k skończona i wtedy $N(E) \geq 2^{n-1}$ i jest ok albo E jest p_k -nieskończona dla wszystkich k .

W drugim przypadku dla każdego k wybieramy minimalne p_k -nieskończone podwyrażenie F_k^* w wśród tych F_k^* wybieramy minimalne względem relacji zawierania i nazywamy je F^* . F^* ma punkt bazowy, powiedzmy j . wybieramy G^* spośród F_k^* pokrywających p_{j+1} .

Teraz jeśli w E zastąpimy F^* przez ϵ , G^* nadal pokrywa p_{j+1} po zamianie (jako że p_{j+1} omija j). Więc

$$N(G^*) \geq 2^{n-2} \text{ po podstawieniu}$$

$$N(F^*) \geq 2^{n-2} \text{ przed podstawieniem}$$

$$N(E) \geq 2^{n-1} \text{ przed podstawieniem}$$

Ponieważ mamy dowieść, że dowolne wyrażenie regularne opisujące język $L(A_n)$ ma długość co najmniej c^n , więc bierzemy stałą $c>1$ i mniejszą od 2, np. 1.5. Wtedy $(1.5)^n \leq 2^n \leq N$ dla $n > 2$. c.n.d.

zadanie zaczerpnięte artykułu: Complexity measures for regular expressions - Andrzej Ehrenfeucht, Paul Zeiger