Лема про розростання

У теорії формальних мов велике значення мають твердження, в яких формулюється необхідна умова приналежності мови до того чи іншого класу мов. Ці твердження відомі в літературі за назвою лем про розростання (або лем про “накачування”). За допомогою цих лем вдається довести, що та чи інша мова не є мовою цього класу, наприклад, не є регулярною, не є контекстно-вільною тощо. Доводити такі “негативні” твердження набагато важче, ніж “позитивні” (що мова є мовою цього класу), бо в останньому випадку достатньо придумати будь-яку граматику відповідного класу, яка породжує цю мову, тоді як в першому потрібно якось довести, що не існує граматики цього класу, яка породжує мову.

Застосування лем про розростання полягає в такому: довівши, що мова не задовольняє умову леми про розростання, ми можемо бути впевнені в тому, що вона не належить до відповідного класу мов.

Лема про розростання для регулярних мов

У цій лемі стверджується, що будь-яка регулярна мова допускає представлення всіх своїх ланцюжків у вигляді з'єднання трьох ланцюжків, причому середній ланцюжок з цих трьох не є порожнім, обмеженим за довжиною, і його "накачка" – повторення будь-яку кількість разів – або викидання НЕ ВИВОДИТЬ за межі мови (тобто дає ланцюжки, що належать цій регулярній мові).

Якщо L – регулярна мова, то існує натуральна константа k_L (залежна від L), така, що для будь-якого ланцюжка $x \in L$, довжина якого не менша за k_L, x допускає представлення у вигляді $x = uvw$, де $v \neq \lambda$ і $|v| \leq k_L$, причому для будь-якого $n \geq 0$ ланцюжок $x_n = u\,v^n\,w \in L$.

Доведення регулярності або нерегулярності мови. **Приклад:**

Доведемо нерегулярність мови
$L(M) = \{ a^n b^n, n \geq 0 \}$.

Вибираючи n настільки великим, щоб воно перевищувало k_L (константу леми), одержуємо такі можливі випадки розміщення середнього v в ланцюжку $a^n b^n$. Зокрема можливі варіанти:

$$\underbrace{a \ldots \overbrace{aa \ldots a}^{v=a^s}}_{n\ \text{раз}}\ \underbrace{b \ldots b}_{n\ \text{раз}}$$

1. $v = a^s,\ s < n$, тобто "накачуваний" підланцюжок v цілком розташовується в "зоні символів a".

Накачування в цьому випадку виведе за межі мови, оскільки при повторенні ланцюжка v кількість символів a необмежено зростатиме, а кількість символів b залишатиметься сталою.

$$\underbrace{a \ldots a}_{n\ \text{раз}}\ \underbrace{\overbrace{bb \ldots b}^{v=b^s}\ b..b}_{n\ \text{раз}}$$

2. $v = b^s, s < n$, тобто "накачуваний" підланцюжок v цілком розташовується в "зоні символів b".

Накачування неможливе з тієї ж причини, що і в попередньому випадку.

$$\underbrace{a \ldots \overbrace{aa \ldots a}^{a^p}}_{n\ \text{раз}}\ \underbrace{\overbrace{b \ldots b}^{b^q} \ldots b}_{n\ \text{раз}}$$

3. $v = a^p b^q$, де $0 < p < n, 0 < q < n$, тобто "накачуваний" підланцюжок v розташовується на стику зон символів a і b.

У цьому випадку при накачуванні підланцюжок ab входить в слово, яке вже не належить мові L.

Бачимо, що існують ланцюжки, для яких жодні представлення у вигляді з'єднання трьох ланцюжків не задовільняють умови леми про розростання для регулярних мов.

Отже, мова $a^n b^n$ нерегулярна.

Контрольне завдання №21

Довести регулярність мови $L(M) = \{ a^n b^m , n \geq 0 , m \geq 0 \}$.