

Глава 2

Формални езици и граматики. Крайни автомати без памет

2.1 Думи и езици

Нека V е крайно и непразно множество. Елементите на това множество ще наричаме *букви*, а самото множество V – *азбука*. По нататък ще разглеждаме азбуки, за които $|V| \geq 2$.

Определение 2.1.1 Дума над азбуката V ще наричаме всяка крайна редица от букви на V .

Една и съща буква може да присъствува няколко пъти в думата.

Дума, в която не влиза нито една буква се нарича *празна дума* и се бележи със символа ε .

Например, ако $V = \{0, 1\}$, то ε , 0, 0010 са думи над азбуката V , ако $V = \{if, then, else, do, a, b\}$, то $\alpha = if\ b\ then\ do\ a\ else\ do\ a$ е дума над тази азбука.

Под *дължина* на една дума α се разбира броят на буквите в нея и се означава с $|\alpha|$.

Например, ако $\alpha = 0100$, то $|\alpha| = 4$.

В съвкупността от думите над дадена азбука се въвежда операцията конкатенация.

Нека α и β са две думи над азбуката V . Под *конкатенация* $\alpha\beta$ на двете думи ще разбираме думата, получена от последователното дописване на буквите на β след последната буква на α .

Например, ако $\alpha = 0110$ и $\beta = 111$, то $\alpha\beta = 0110111$.

Под α^n ще разбираме конкатенацията $\underbrace{\alpha\alpha\ldots\alpha}_{n\text{-пъти}}$ и $\alpha^0 = \varepsilon$.

Да означим с V^* множеството от всички думи над азбуката V . Операцията конкатенация притежава следните свойства:

1. $\alpha(\beta\gamma) = (\alpha\beta)\gamma$; 2. $\varepsilon\alpha = \alpha\varepsilon = \alpha$

за всеки три думи α, β, γ от V^* . Следователно V^* е моноид с единица ε .

Казваме, че думата α е *начало* (*префикс*) на думата β , ако съществува дума γ такава, че $\beta = \alpha\gamma$. Ако съществува дума δ така, че $\beta = \delta\alpha$, то α се нарича *край* (*суфикс*) на β . Когато могат да се намерят думи γ и δ така, че $\beta = \gamma\alpha\delta$, то казваме, че α е *поддума* на β .

Очевидно в множеството V^* , операцията конкатенация е вътрешна (алгебрична) т.е. за всеки две думи $\alpha, \beta \in V^*$ думата $\alpha\beta$ е от V^* .

Оказва се, че всички думи в V^* могат да бъдат номерирани, откъдето следва, че множеството V^* е изброимо. Тази номерация може да се осъществи например, чрез следното изображение $\text{Nom} : V^* \rightarrow N$ дефинирано с рекурентното съотношение

$$\text{Nom}(\alpha) = \begin{cases} m & \text{ако } \alpha = a_m, a_m \in V \\ n.\text{Nom}(\beta) + \text{Nom}(a), & \text{ако } \alpha = \beta a, a \in V, \beta \in V^* \end{cases},$$

където $V = \{a_1, a_2, \dots, a_n\}$ и $\text{Nom}(\varepsilon) = 0$.

Теорема 2.1.1 *Функцията Nom е биекция.*

Доказателство: Нека $\text{Nom}(\alpha) \neq \text{Nom}(\beta)$. Да допуснем, че $\alpha = \beta$. Ако $\alpha = a_s \in V$, то $\text{Nom}(\alpha) = s$ и следователно $\beta = a_s$ и $\text{Nom}(\beta) = s$, което е противоречие.

Следователно, ако $|\alpha| = 1$ и $\text{Nom}(\alpha) \neq \text{Nom}(\beta)$, то $\alpha \neq \beta$.

Да допуснем, че при $|\alpha| \leq k$, ако $\text{Nom}(\alpha) \neq \text{Nom}(\beta)$, то $\alpha \neq \beta$.

Нека сега $|\alpha| = k + 1$ и $\text{Nom}(\alpha) \neq \text{Nom}(\beta)$, а да приемем, че $\alpha = \beta$, като $\alpha = \gamma a_{i_{k+1}}$ и $\beta = \delta a_{i_{k+1}}$, $a_{i_{k+1}} \in V$.

Тогава

$$\text{Nom}(\alpha) = n\text{Nom}\gamma + i_{k+1} \quad \text{и} \quad \text{Nom}(\beta) = n\text{Nom}\delta + i_{k+1}.$$

Следователно $\text{Nom}(\gamma) \neq \text{Nom}(\delta)$, но $\gamma = \delta$. Тъй като $|\gamma| = k$, то последното равенство е в противоречие с индуктивното допускане.

С това доказахме, че Nom е изображение.

Да предположим, че $\text{Nom}(\alpha) = \text{Nom}(\beta)$.

Очевидно, ако α е буква, то $\text{Nom}(\alpha) \leq n$ и следователно β също е буква от V , но тогава $\alpha = \beta$.

Нека $|\alpha| \geq 2$ и $|\beta| \geq 2$. Тогава съществуват единствени числа q и r така, че

$$\text{Nom}(\alpha) = \text{Nom}(\beta) = nq + r \text{ и } 0 \leq r < n.$$

Съгласно дефиницията на Nom следва, че $\alpha = \gamma a$ и $\beta = \delta b$, където $a, b \in V$ и $\gamma, \delta \in V^*$, като $\text{Nom}(a) = \text{Nom}(b) = r$ и $\text{Nom}(\gamma) = \text{Nom}(\delta) = q$. Понеже $|\gamma| < |\alpha|$ и $|\delta| < |\beta|$ индуктивно следва, че $\gamma = \delta$ и $a = b$. Следователно $\alpha = \beta$, което означава, че Nom е влягане.

Да докажем сега, че Nom е сюрективно изображение.

Нека r е произволно естествено число. Ако $r = 0$, то $r = \text{Nom}(\varepsilon)$, а ако $r \leq n$, то $r = \text{Nom}(a_r)$, $a_r \in V$.

Да допуснем, че за всяко $r \leq k$ съществува дума α от V^* така, че $r = \text{Nom}(\alpha)$.

Да докажем, че съществува дума $\alpha \in V^*$ така, че $k + 1 = \text{Nom}(\alpha)$. Разглеждаме двата възможни случая:

(i) $k + 1$ не е кратно на n , т.е. съществуват числа q и r така, че $k + 1 = nq + r$, $0 < r < n$.

Очевидно $q \leq k$ и съгласно индуктивното допускане съществува дума β , така че $q = \text{Nom}(\beta)$.

Лесно се вижда, че

$$k + 1 = \text{Nom}(\beta a_r), \quad a_r \in V.$$

(ii) $k + 1$ е кратно на n , т.е. $k + 1 = qn$, $q \leq k$.

Ако $q = 1$, то $k + 1 = \text{Nom}(a_n)$, $a_n \in V$.

Ако $q > 1$, то нека β е такава дума, че $q - 1 = \text{Nom}(\beta)$. В този случай $k + 1 = \text{Nom}(\beta a_n)$, $a_n \in V$.

Следователно Nom е биективно изображение. \square

Пример 2.1.1 Нека $V = \{0, 1, a\}$. Тогава $\text{Nom}(\varepsilon) = 0$, $\text{Nom}(0) = 1$, $\text{Nom}(1) = 2$, $\text{Nom}(a) = 3$. Да пресметнем $\text{Nom}(\alpha)$, ако $\alpha = a010$.

Последователно получаваме

$$\begin{aligned} \text{Nom}(\alpha) &= 3\text{Nom}(a01) + 1 = 3(3\text{Nom}(a0) + 2) + 1 = \\ &= 3.(3.(3\text{Nom}(a) + 1) + 2) + 1 = 3.(3.(3.3 + 1) + 2) + 1 = 97. \end{aligned}$$

Определение 2.1.2 Нека V е азбука. Всяко подмножество L на V^* се нарича **формален език (или само език)** над азбуката V .

В частност V^* е език над V , а също така \emptyset е език над азбуката V , който се нарича *празен език*. Да отбележим, че празният език се различава от езика $L = \{\varepsilon\}$, който се състои само от празната дума ε .

Нека L_1 и L_2 са два езика над азбуката V . Под *сечение (обединение)* на двата езика ще разбираме езика $L_1 \cap L_2$ ($L_1 \cup L_2$), който се получава от сечението (обединението) на L_1 и L_2 разглеждани като множества от думи над V .

Под *произведение на езиците* L_1 и L_2 ще разбираме езика $L_1 L_2 = \{\alpha \in V^* \mid \alpha = \beta\gamma \text{ и } \beta \in L_1 \text{ и } \gamma \in L_2\}$, т.е. множеството от всички думи, които се представят като конкатенации между думи на L_1 и L_2 .

Очевидно в общия случай

$$L_1 L_2 \neq L_2 L_1.$$

Поради асоциативното свойство на конкатенацията, обаче може да се докаже, че за всеки три езика L_1, L_2 и L_3 е в сила

$$L_1(L_2 L_3) = (L_1 L_2)L_3.$$

Под n -та степен на езика L ще разбираме езика

$$L^n = \underbrace{LL \dots L}_{n\text{-пъти}} \text{ като } L^0 = \{\varepsilon\}.$$

Езикът $L^* = \bigcup_{n \geq 0} L^n$ се нарича *итерация* на L , а $L^+ = \bigcup_{n > 0} L^n$ – *положителна итерация* на L .

Пример 2.1.2 Нека $V = \{0, 1, a, b\}$, $L_1 = \{\varepsilon, 01, a1\}$, $L_2 = \{\varepsilon, 02, 1b\}$. Тогава

$$L_1 \cap L_2 = \{\varepsilon\},$$

$$L_1 \cup L_2 = \{\varepsilon, 01, a1, 02, 1b\},$$

$$L_1 L_2 = \{\varepsilon, 01, a1, 02, 1b, 0102, 011b, a102, a11b\}.$$

Нека да разгледаме азбуките V и W и съответните им множества V^* и W^* от думи над тях.

Ще казваме, че изображението $h : V^* \rightarrow W^*$ е *хомоморфизъм*, ако за всеки две думи α и β от V^* е в сила

$$h(\alpha\beta) = h(\alpha)h(\beta).$$

Когато h е взаимно-еднозначно изображение, то се нарича *изоморфизъм* между моноидите V^* и W^* .

Хомоморфизмите се оказват един доста прост начин за генериране на езици над дадена азбука.

Нека V е азбука и $h : V^* \rightarrow V^*$ е хомоморфизъм на V^* в себе си. За произволна дума α над азбуката V , можем да разгледаме езика L , определен по следният начин

$$L = L(h, \alpha) = \{\beta \in V^* \mid \beta = h^i(\alpha), i \geq 0\},$$

където

$$h^i(\alpha) = \underbrace{h(h(\dots(h(\alpha)\dots)))}_{i\text{-пъти}} \text{ като } h^0(\alpha) = \alpha.$$

Езиците, породени по този начин имат приложение при разглеждане на някои задачи за разрешимост, а също така и за установяване на еквивалентност между езици. Някои от свойствата на тези езици се свързват с модели на саморазвиващи се организми в биологията.

З а д а ч и

1. Ако $L_1 = \{ab, ca\}$ и $L_2 = \{ca, abc\}$ са два формални езика над азбуката $V = \{a, b, c\}$, да се опишат езиците

$$L_1 \cup L_2, L_1 \cap L_2, L_1 L_2, L_1^2 \cup L_2^2.$$

2. Да се докаже, че за всеки три формални езика L_1, L_2 и L_3 над дадена азбука V е в сила $L_1(L_2 L_3) = (L_1 L_2)L_3$.

3. Нека u е най-кратката поддума на думата v , за която $v = u^n$ за някое естествено n . Думата u в този случай да означим с $p(v)$.

а) да се докаже, че ако $p(v)$ съществува, то тя е единствена;

б) да се докаже, че $vw = wv$ е в сила точно тогава, когато $p(w) = p(v)$;

в) да се докаже, че ако $z^m v^n = w^t$ за някои цели числа $m, n, t \geq 2$, то $p(z) = p(v) = p(w)$.

4. Нека $V = \{a, b\}$ е азбука и $h : V^* \rightarrow V^*$ е хомоморфизъм, дефиниран по следния начин

$$h(a) = b \text{ и } h(b) = ab.$$

Да се докаже, че дължините на думите в езика $L(h, a) = \{h^i(a) \mid i \geq 1\}$ образуват редицата от числата на Фибоначи.

5. Да се докаже, че всеки хоморфизъм h между V_1^* и V_2^* се определя напълно от действието му върху азбуката V_1 , т.е. ако са зададени само образите при h на еднобуквените думи във V_1^* .

6. Да се докаже, че един хоморфизъм между V_1^* и V_2^* е изоморфизъм точно тогава, когато за всяка буква $a \in V_1$

$$|h(a)| = 1 \text{ и } |V_1| = |V_2|.$$