1.6. ПОДРЕДЕНИ МНОЖЕСТВА. ОПЕРАЦИИ С НИЗОВЕ. РЕГУЛЯРНИ МНОЖЕСТВА И ИЗРАЗИ

1. Подредени множества и низове

Определение:

Множества, чиито елементи (низове или думи) са подредени в ред, при което са дефинирани първи последен, предходен, следващ елемент и т.н. се наричат подредени множества.

Произволно крайно множество от символи се нарича азбука, която се означава $\sum=\{a,b\}$, където a,b са символи или букви на азбуката. Всяка крайна редица от букви от азбуката се нарича низ или дума над Σ.

Пример:

Символни низове (думи) над $\sum=\{a,b\}$: $u=aabaa, v=bbbabba$.
Символни низове над $\sum=\{0,1\}$: $001,\ 0011$.
Дума над $\sum=\{if, then, else, do, a, b\}$: $\alpha = if\ b\,then\,do\,a\,else\,do\,a$.
Броят на буквите в една дума α дефинира дължината на думата, означава се с $|\alpha|$.
Пример:
Ако думата е $\alpha=110011$, то дължината на думата е $|\alpha|=6$.

Конкатенация на думите α и β над $\sum=\{a,b\}$ е думата $\alpha\beta$, получена от последователното дописване на буквите на β след последната буква на α.
Пример:

Ако $\alpha=1101$ и $\beta=110$, то $\alpha\beta=1101110$.

Под α^n се разбира конкатенацията $\underbrace{\alpha.\alpha...\alpha}_{n-\text{пъти}}$, а $\alpha^0=\varepsilon$, т.е. празната буква.

Свойства на конкатенацията на думите:

Ако α,β,γ са три думи над Σ* в сила са следните съотношения

1. $\alpha(\beta\gamma)=(\alpha\beta)\gamma$.
2. $\varepsilon\alpha=\alpha\varepsilon=\alpha$.
3. Думата α е начало (префикс) на думата β, ако съществува дума γ такава, че $\beta=\alpha\gamma$. Думата α е край (суфикс) на думата β, ако съществува дума δ такава, че $\beta=\delta\alpha$. Думата α е поддума на β, ако съществуват думи γ и δ така, че $\beta=\gamma\alpha\delta$.

Означения:

Λ - празна дума - думата, която не съдържа букви;

Σ* - множество от всички думи над Σ, включително и празната дума Λ, ако $\sum\{a,b\}$, $\sum^* = \{\Lambda, a, b, aa, ab, ba, bb, aaa, ...\}$;
$\{\ \}$, $\varnothing$ - празно множество, т.е. множество от думи, което не съдържа нито една дума;
$\{\Lambda\}$ - непразно множество, състоящо се от една единствена дума – празната дума Λ.

2. Операции с множества от низове

Конкатенация (съединяване или умножаване) UV на две подмножества U, V над Σ* се дефинира с израза

$$UV = \{x | x = uv, u \in U, v \in V\}.$$

Свойства:

Ако $U = \{a, ab, aab\}$ и $V = \{b, bb\}$ то конкатенацията UV е $UV = \{ab, abb, abbb, aabb, aabbb\}$, а конкатенацията VU е $VU = \{ba, bab, bab, bba, bbab, bbaab\}$. Конкатенацията не е комутативна ($UV \neq VU$). Конкатенацията е асоциативна, т.е. за произволни подмножества U, V и W на Σ* е в сила $(UV)W = U(VW)$.

Затворена обвивка (звезда) на множеството S е множеството S^*, състоящо се от празната дума и всички думи, образувани чрез конкатенация на краен брой думи на S.

Пример:
Ако $S = \{ab, bb\}$, то $S^* = \{\Lambda, ab, bb, abab, abbb, abbb, bbab, bbbb, ababab....\}$.

Звезда, дефинирана с обединения на подмножествата S^i за $i > 0$:

$$S^* = S^0 \cup S^1 \cup S^2 \cup S^3 \cup ...., S^i,$$

където $S^0 = \{\varepsilon\}$ или $S^0 = \{\Lambda\}$ е празната буква, $S^i = S^{i-1}S$ за $i > 0$ - конкатенацията има вида $S^i = S.S\ldots.S$ (i-пъти).

3. Регулярни множества

Класът на регулярните множества над Σ се дефинират рекурсивно в съответствие с правилата:

Всяко крайно множество от думи над Σ* (включително празното множество ∅) е регулярно множество.

Ако U и V са регулярни множества над Σ, то $U \cup V$ и UV са регулярни.

Ако S е регулярно множество над Σ, то S^* е също регулярно множество.

Множеството е регулярно само, ако се получава чрез краен брой прилагания на правилата от 1 до 3.

Класът от крайни множества над Σ, е най-малкият клас, който съдържа всички крайни множества от думи над Σ и е затворен относно операциите: обединение, конкатенация и звезда.

Множество от всички думи над $\sum = \{a,b\}$, съдържащи *aa* и/или *bb*, както и множеството от всички думи над $\sum = \{a,b\}$, съдържащи четен брой *a* и четен брой *b*, са регулярни над Σ: $T = \{aa, aab, abb, abba, baab, ...\}$.

Множеството $T = \{a^n b^n | n \geq 0\}$ от всички думи, които се състоят от *n* “*a*”, последвани от *n* “*b*”, при $n \geq 0$, не е регулярно множество над Σ.

Дефиниции:

- дума: *bbaba*;

множество от думи: $\{a^n b^n | n \geq 0\}$;

клас от множество от думи:

$$\{\{a^n | n \geq 0\}, \{a^n b^n | n \geq 0\}, \{a^n b^n a^n | n \geq 0\}, \{a^n b^n a^n b^n | n \geq 0\}, ...\}.$$

4. Регулярни изрази

Всеки регулярен израз *R* над Σ описва множество $\widetilde{R}$ от думи над Σ, т.е. $\widetilde{R} \subseteq$ Σ*, което се дефинира рекурсивно по следния начин:

Ако $R = \Lambda$, то $\widetilde{R} = \{\Lambda\}$ - множество, състоящо се от празната дума, ако *R*=∅, то $\widetilde{R}$ = {∅} – празно множество.

Ако *R* = σ, то $\widetilde{R}$ = {σ} – множество, състоящо се от буквата σ.

Ако R_1 и R_2 са регулярни изрази над Σ, описващи множество от думи $\widetilde{R}_1$ и $\widetilde{R}_2$:

при $R = (R_1 + R_2)$, то $\widetilde{R} = \widetilde{R}_1 \cup \widetilde{R}_2 = \{x | x \in \widetilde{R}_1, x \in \widetilde{R}_2\}$ дефинира обединение на двете множества;

при $R = (R_1 . R_2)$, то $\widetilde{R} = \widetilde{R}_1 . \widetilde{R}_2 = \{xy | x \in \widetilde{R}_1, y \in \widetilde{R}_2\}$ дефинира произведение на двете множества;

при $R = (R_1)^*$, то $\widetilde{R} = \widetilde{R}_1^* = \{\Lambda\}$ ∪ { $x | x$ е получено чрез съединяване на краен брой думи от $\widetilde{R}_1$ } е затворената обвивка на множеството $\widetilde{R}_1$.

Взаимно еднозначното съответствие на регулярните изрази над $\sum = \{a,b\}$ и регулярните множества може да се илюстрира със следните примери:

$R = ba^* \to \widetilde{R}$ = Всички думи над Σ, започващи с b, последвано само от a-та.

$R = a^*ba^*ba^* \to \widetilde{R}$ = Всички думи над Σ, съдържащи точно две b-та;
$R = (a+b)^* \to \widetilde{R}$ = Всички думи над Σ.
$R = (a+b)^*(aa+bb)(a+b^*) \to \widetilde{R}$ = Всички думи над Σ, съдържащи две последователни a-та и две последователни b-та.
$R = [aa+bb+(ab+ba)^*(aa+bb)(ab+ba)]^* \to \widetilde{R}$ = Всички думи над Σ, съдържащи четен брой a-та и четен брой b-та.
$R = (b+abb)^* \to \widetilde{R}$ = Всички думи над Σ, в които всяко a е непосредствено последвано поне от две b-та.

Определение: Едно множество е регулярно над Σ точно тогава, когато може да се представи чрез регулярен израз над Σ. Два регулярни израза R_1 и R_2 над Σ са еквивалентни (означава се: $R_1 = R_2$) точно тогава, когато $\widetilde{R}_1 = \widetilde{R}_2$.

За произволни регулярни изрази R, S и T са в сила следните формули за еквивалентност:

1. $R + S = S + R$, $R + \varnothing = \varnothing + R$, $R + R = R$,
$(R + S) + T = R + (S + T)$.
2. $R\,\Lambda = \Lambda\,R = R$, $R\,\varnothing = \varnothing\,R = \varnothing$, $(R\,S)\,T = R\,(S\,T)$.
В общия случай: $R\,S \neq R\,S$.
3. $R\,(S + T) = R\,S + R\,T$, $(S + T)\,R = S\,R + T\,R$.
$R^* = R^*\,R^* = (R^*)^* = (\Lambda + R)^*$, $\varnothing^* = \Lambda^* = \Lambda$.
$R^* = \Lambda + R + R^2 + \ldots + R^k + R^{k+1}\,R^*$ $(k \geq 0)$
В частност $R^* = \Lambda + R\,R^*$.
$(R + S)^* = (R^* + S^*)^* = (R^*\,S^*)^* = (R^*\,S)\,R^* = R^*\,(S\,R^*)^*$.

В общия случай $(R + S)^* \neq R^* + S^*$

$R^*\,R = R\,R^*$, $R\,(S\,R)^* = (R\,S)^*R$.
$(R^*\,S)^* = \Lambda + (R + S)^*S$, $(R\,S^*)^* = \Lambda + R(R + S)^*$.
Правило на Ардън. Допуска се, $\Lambda \notin S$. Тогава:

$R = S\,R + T$ точно тогава, когато $R = S^*\,T$;
$R = R\,S + T$ точно тогава, когато $R = T\,S^*$.