Формални Езици и Езикови Процесори
ТУ, кат. КС, летен семестър 2012

Лекция 3

Тема:

Регулярни Изрази

Съдържание:

- Регулярни множества
- Регулярни изрази
- Примери:
  - Цели десетични константи без знак
  - Цели десетични константи със знак
  - Идентификатори
  - Осмични, 16-чни, двоични константи
  - Реални десетични константи

Регулярни множества

Операции над множества от низове – припомняне от л.1-2

Наред с познатите операции над множества (обединение, сечение, разлика, допълнение), се въвеждат три специфични операции над множества от символни низове.

Нека означим като операнди P и Q две множества от низове.

- Дизюнкция (Alternation);
- Конкатенация (Concatenation);
- Итерация (Kleene closure or full iteration);
- Непразна итерация (Positive closure or non-empty iteration);

Операции над множества от символни низове

- Дизюнкция
- Конкатенация
- Итерация

Операции над множества от символни низове

- Дизюнкция (двуместна /бинарна/ операция с операнди P и Q)
  – Определение: обединение на две множества

$$Z = P \mid Q$$

  – Примери

Операции над множества от символни низове

- Конкатенация (двуместна /бинарна/ операция с операнди P и Q)
  - Определение

$$Z = P \,.\, Q = P\,Q$$

  - Примери

P={0,1} Q={a,b} PQ={0a, 0b, 1a, 1b}

Операции над множества от символни низове

- Итерация (едноместна /унарна/ операция с операнд P)

  – Определение

$$Z = P^* = \{\varepsilon\} \mid P \mid P.P \mid P.P.P \mid \ldots$$

  – Примери

Операции над множества от символни низове

- Непразна итерация (едноместна /унарна/ операция с операнд P)
  - Определение

$$Z = P^{+} = P \mid P.P \mid P.P.P \mid \ldots$$

  - Примери

Операции над множества от символни низове

- Релация: итерация – непразна итерация

$$P^* \quad = \{\varepsilon\} \mid P \mid P.P \mid P.P.P \mid \ldots$$

$$P^+ \quad = P \mid P.P \mid P.P.P \mid \ldots$$

$$P^+ = P \,.\, P^* = P \;\; P^*$$

Операции над множества от символни низове

- Приоритет на операциите

висок

↑ Итерация - Iteration

↑ Конкатенация - Concatenation

↑ Дизюнкция - Alternation

нисък

Операции над множества от символни низове

Не се налагат ограничения над операндите като множества от низове. Изброените операции са приложими над всякакви множества от символни низове.

Естествен подход за формализиране на проблема е да се възприеме операндите на операциите да са едноелементни множества от символни низове, т.е. множества с един низ и мощност 1.

Такъв тип множества се определя като ***базови множества*** или още ***регулярни множества***.

Формално определение на РМ

Едно множество от символни низове е регулярно множество над азбуката Σ, ако е определено рекурсивно по следния начин:

1. $\varnothing$ (празно множество) е регулярно множество;
2. $\{\varepsilon\}$ е регулярно множество;
3. $\{a\}$ е регулярно множество за произволно $a \in \Sigma$;
4. Ако P и Q са регулярни множества от низове над азбуката Σ, такива са и множествата, получени чрез прилагане на операциите конкатенация, дизюнкция и итерация:
   - Конкатенацията $P.Q$ (или само PQ) е равна на $\{xy \mid x \in P$ и $y \in Q\}$;
   - Дизюнкцията на множествата P и Q е множеството, всеки елемент на което принадлежи или на P, или на Q, т.е. $P \mid Q = \{x \mid x \in P$ или $x \in Q\}$;
   - Итерацията на едно множество P се означава като P^* и представлява дизюнкция на множествата $\{\varepsilon\}$, P, PP, PPP, …. Итерацията се определя по следния начин: $P^* = \mid P$n, където $P_n = PP_{n-1}$ и $P_0 = \{\varepsilon\}$; n≥0

1. Нищо друго не е регулярно множество от низове над азбука Σ.

Регулярни изрази

Регулярни изрази

Регулярните изрази са удобен начин за задаване на регулярни множества. Те се образуват от операнди, представляващи множества от низове, и знаците за операции (подредени според естествения приоритетен ред):

- итерация (*)
- конкатенация (.)
- дизюнкция (|)

Приоритетният ред може да се промени чрез въвеждане на скоби ( ) за формиране на под изрази.

Регулярни изрази

РИ са средство за задаване, описание, определение на РМ.

Формалната дефиниция на РИ включват 5 правила, които съответстват на 5-те правила на определение на РМ.

Регулярни изрази

Изразите се определят рекурсивно по следния начин:

1. $\varnothing$ е регулярен израз, означаващ регулярното множество $\varnothing$;
2. ε е регулярен израз, означаващ регулярното множество $\{\varepsilon\}$;
3. Ако $a \in \Sigma$, a е регулярен израз, означаващ множеството $\{a\}$;
4. Ако p и q са регулярни изрази, означаващи регулярните множества P и Q, тогава:
   – $(p|q)$ е регулярен израз, означаващ $P|Q$;
   – (pq) е регулярен израз, означаващ $P.Q$;
   – $(p)^*$ е регулярен израз, означаващ P^*;

1. Нищо друго не е регулярен израз.

С $p+$ съкратено се означава изразът pp^*.

Съответствие РМ - РИ

а/ $\varnothing$ (празно множество) е регулярно множество;

а/ $\varnothing$ е регулярен израз, означаващ регулярното множество $\varnothing$;

б/ $\{\varepsilon\}$ е регулярно множество;

б/ ε е регулярен израз, означаващ регулярното множество $\{\varepsilon\}$;

в/ $\{a\}$ е регулярно множество за произволно $a \in \Sigma$;

в/ Ако $a \in \Sigma$, a е регулярен израз, означаващ множеството $\{a\}$;

Съответствие РМ - РИ

г/ Ако P и Q са регулярни множества от низове над азбуката Σ, такива са и множествата, получени чрез прилагане на операциите конкатенация, дизюнкция и итерация:

- Конкатенацията $P.Q$ (или само PQ) е равна на $\{xy \mid x\in P$ и $y\in Q\}$;
- Дизюнкцията на множествата P и Q е множество, всеки елемент на което принадлежи или на P, или на Q, $P|Q = \{x \mid x \in P$ или $x \in Q\}$;
- Итерацията на едно множество P се означава като P^* и представлява дизюнкция на множествата $\{\varepsilon\}$, P, PP, PPP, ...

г/ Ако p и q са регулярни изрази, означаващи регулярните множества P и Q, тогава:

– $(p|q)$ е регулярен израз, означаващ $P \mid Q$;

– (pq) е регулярен израз, означаващ $P.Q$;

– $(p)^*$ е регулярен израз, означаващ P^*;

Съответствие РМ - РИ

д/ Нищо друго не е регулярно множество над Σ
д/ Нищо друго не е регулярен израз.

РИ - примери

Следват 4 примера на РИ, които служат за описание/представяне/ на безкрайни РМ.

- Изразът $(a|b)^*$ задава множеството от низове $\{a, b\}^*$, чиито елементи са: ε, *a*, *b*, *aa*, *ab*, *aaa*, *aab*, *baa*, *bab*, *aaab*,…

РИ - примери

Изразът $a(b|c)^*$ определя множеството на всички низове, започващи с *a*, след което може да следва произволна последователност от символите *b* или *c*, т.е. $\{a\}\{b, c\}^*$. Елементи на множеството са например низовете: *a*, *ab*, *ac*, *abb*, *abc*, *acb*, *acc*, *abbb*, ...

РИ - примери

- Изразът $a^*(b|c)$ определя множеството на всички низове, започващи с произволна последователност от символи a, след което завършва със символ b или символ c, т.е. $\{a\}^*\{b, c\}$, чиито елементи са: b, c, ab, ac, aab, aac, ...

РИ - примери

- Изразът 01(0|1)* задава множеството {01}{0, 1}* от всичките низове, съдържащи 0 и 1 и започващи с 01.

РИ - примери

a

a|b

ab

a(b|c)

(aa|b)*

a*

a+

a.b* = ab*

a*.b = a*b

a*b*

РИ – свойства и еквивалентни преобразования

Два регулярни израза се свързват със знак за равенство (=), ако те означават едно и също множество.

Съществува лема, установяваща някои основни алгебрични свойства на регулярните изрази (p, q и r са произволни регулярни изрази):

Следните тъждества представят равенство на РИ и служат за еквивалентни преобразования на РИ.

РИ – свойства и еквивалентни преобразования

1. $p|q = q|p$ комутативно правило
2. $p|(q|r) = (p|q)|r$ асоциативно правило
3. $p(q|r) = pq|pr$ лява факторизация
4. $(p|q)r = pr|qr$ дясна факторизация
5. $p\varepsilon = \varepsilon p = p$
6. $p|p = p$

РИ – свойства и еквивалентни преобразования

1. $p(qr) = (pq)r = pqr$ асоциативно правило
2. $p^*|p = p^*$
3. $(p^*)^* = p^*$
4. $\varnothing^* = \varepsilon$
5. $p|\varnothing = p$
6. $p\varnothing = \varnothing p = \varnothing$

Регулярен Език - РЕ

- – формално описание (определение)

$$\mathrm{RL} = (\Sigma, r)$$

Σ - азбука;

r - регулярен израз.

РИ – съдържателни примери:

– РИ на цели десетични константи без знак;
– РИ на цели десетични константи със знак;
– РИ на идентификатори;
– РИ на реални десетични константи със знак;
– РИ на реални десетични константи със знак в експоненциален запис.

Регулярни Изрази

Обект: цели десетични константи без знак

– Дефиниция (описание)

– Примери

– Регулярен израз

Регулярни Изрази

Обект: цели десетични константи без знак

- Дефиниция (описание)
- Примери
- Регулярен израз

$$R = (0|1|2|\ldots\ |9)\ (0|1|2|\ldots|9)^*$$

$$R = (0\ |\ 1\ |\ 2\ |\ \ldots\ |\ 9)^+ \ = \ \mathrm{D}^+$$

$$D = (\ 0\ |\ 1\ |\ 2\ |\ 3\ |\ \ldots\ |\ 9\ )$$

Регулярни Изрази

Обект: цели десетични константи със знак

– Дефиниция (описание)

– Примери

– Регулярен израз

Регулярни Изрази

Обект: цели десетични константи със знак

- Дефиниция (описание)
- Примери
- Регулярен израз

$$R = (+|–|\varepsilon)\ (0|1|2|\ldots\ |9)\ (0|1|2|\ldots|9)^*$$

$$R = (+|–|\varepsilon)\ (0|1|2|\ldots|9)^+ = (+|–|\varepsilon)D^+$$

$$D = (0|1|2|3|\ldots|9)$$

Регулярни Изрази

Обект: идентификатори

– Дефиниция (описание)

– Примери

– Регулярен израз

Регулярни Изрази

Обект: идентификатори

- Дефиниция (описание)
- Примери
- Регулярен израз

$$R = (A|B|\ldots|Z)\,(A|B|C|\ldots|Z|0|1|2|3|\ldots|9)^*$$

$$R = \mathrm{L}\,(\mathrm{L} \mid \mathrm{D})^*$$

$$L = (\,A \mid B \mid \ldots \mid Z \mid a \mid b \mid \ldots\; z\,)$$
$$D = (\,0 \mid 1 \mid 2 \mid 3 \mid \ldots \mid 9\,)$$

Регулярни Изрази

Обект: Реални десетични константи със знак;

– Дефиниция (описание)

– Примери

– Регулярен израз

Регулярни Изрази

Обект: Реални десетични константи със знак;

- Дефиниция (описание)
- Примери
- Регулярен израз

$$R = (+|-|\varepsilon)\ (D_{10}^{*} \,.\, D_{10}^{+} \mid D_{10}^{+} \,.\, D_{10}^{*})$$

$$D_{10} = (\ 0 \mid 1 \mid 2 \mid 3 \mid \dots \mid 9\ )$$

Регулярни Изрази

Обект: реални десетични константи със знак в експоненциален запис

- Дефиниция (описание)
- Примери
- Регулярен израз

Регулярни Изрази

Обект: осмични константи (C/C++)

– Дефиниция (описание)

– Примери

– Регулярен израз

Регулярни Изрази

Обект: осмични константи (C/C++)

- Дефиниция (описание)
- Примери
- Регулярен израз

$$R = 0\ D_8^{+}$$

$$D_8 = (\ 0 \mid 1 \mid 2 \mid 3 \mid \dots \mid 7\ )$$

Регулярни Изрази

Обект: 16-ични константи (C/C++)

– Дефиниция (описание)

– Примери

– Регулярен израз

Регулярни Изрази

Обект: 16-ични константи (C/C++)

– Дефиниция (описание)
– Примери
– Регулярен израз

$$R = 0\ (X|x)\ D_{16}^{+}$$

$$D_{16} = (0|1|2|3|\ldots|9|A|\ldots|F|a|\ldots|f)$$

Регулярни Изрази

Обект: Двоични константи (Асемблер)

- Дефиниция (описание)
- Примери
- Регулярен израз

Регулярни Изрази

Обект: Двоични константи (Асемблер)

- Дефиниция (описание)
- Примери
- Регулярен израз

$$R = D_2^{+}(B|b)$$

$$D_2 = 0|1$$

Регулярни Изрази

Обект: 16-ични константи (Асемблер)

- Дефиниция (описание)
- Примери
- Регулярен израз

Регулярни Изрази

Обект: 16-ични константи (Асемблер)

- Дефиниция (описание)
- Примери
- Регулярен израз

$$R = D_{10}\, D_{16}^{*}\, (H|h)$$

$$D_{10} = (0|1|2|3|\ldots|9)$$

$$D_{16} = (0|1|2|3|\ldots|9|A|\ldots|F|a|\ldots|f)$$

Графично представяне на РИ

- Синтактични графи – генератори на низове.
- Граф – колекция от възли, някои от които свързани с дъги.
- Насочен граф – дъгите имат посока.

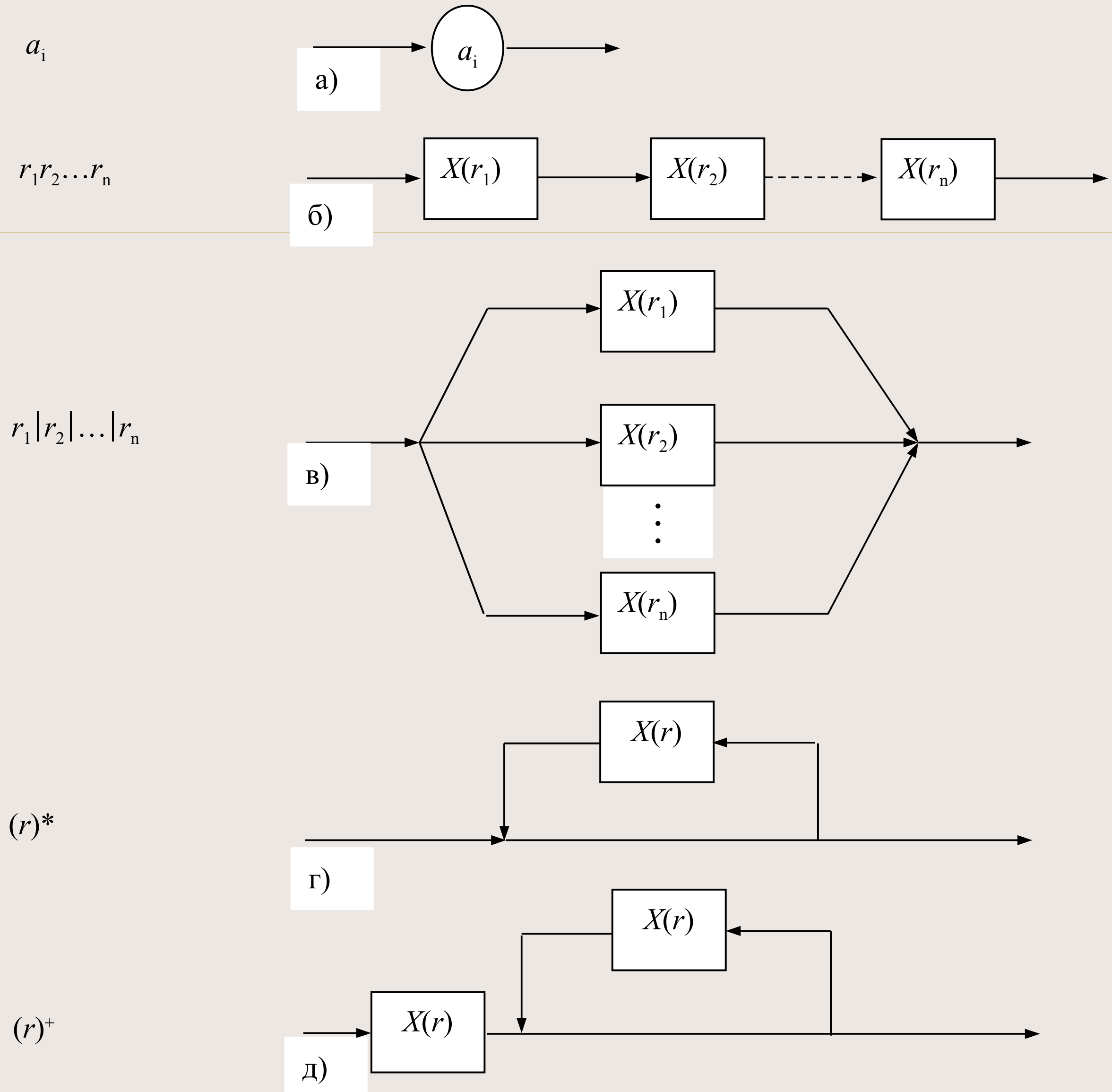
Графично представяне на РИ

- Синтактични графи – генератори на низове.
- Обща структура.
- Три типа възли:
  - Елементарен генератор
  - Възел разклонение
  - Възел обединение

Графично представяне на РИ

Синтактични графи – генератори на низове

- Дизюнкция – илюстрирана
- Конкатенация – илюстрирана
- Итерация – илюстрирана
- Непразна итерация – илюстрирана



Фиг. 2.7. Правила за съставяне на граф-схеми

Синтактични графи - примери

- Цели десетични константи без знак
- Цели десетични константи със знак
- Идентификатори
- Реални десетични константи със знак
- Реални десетични константи със знак в експоненциален запис

Цели десетични константи без знак

$R = (0|1|2|\ldots\ |9)\ (0|1|2|\ldots|9)^*$

$R = (0|1|2|\ldots|9)^+\ =\ \mathrm{D}^+$

$D = (0|1|2|3|\ldots|9)$

Цели десетични константи със знак

$R = (+|–|\varepsilon)\ (0|1|2|\ldots|9)\ (0|1|2|\ldots|9)^*$

$R = (+|–|\varepsilon)\ (0|1|2|\ldots|9)^+ = (+|–|\varepsilon)D^+$

$D = (0|1|2|3|\ldots|9)$

Примери: Идентификатори

$R = (A|B|\ldots|Z)\ (A|B|C|\ldots|Z|0|1|2|3|\ldots|9)^*$

$$R = \mathrm{L}\ (\mathrm{L} \mid \mathrm{D})^*$$

$L = (A|B|\ldots|Z|a|b|...z)$

$D = (0|1|2|3|\ldots|9)$

Примери: Осмични константи

$R = 0\ D_8^{+}$

$D_8 = (0|1|2|3|\ldots|7)$

Примери: 16-ични константи

$R = 0\ (X|x)\ D_{16}^{+}$

$D_{16} = (0|1|2|3|\ldots|9|A|\ldots|F|a|\ldots|f)$

Примери: Двоични константи

$R = D_2^+(B|b)$

$D_2 = 0|1$

Примери:
Реални десетични константи

$R = (+|–|\varepsilon)\ (D_{10}^{*}\ .\ D_{10}^{+}\ |\ D_{10}^{+}\ .\ D_{10}^{*})$

$R = (+|–|\varepsilon)\ (D_{10}^{*}\ .\ D_{10}^{+}\ |\ D_{10}^{+}\ .\ D_{10}^{*})$

$D_{10} = (0|1|2|3|\ldots|9)$

Благодаря

За

Вниманието