Проектирование ДКА, допускающего класс «восьмеричных констант» языка СИ

1. Σ=ASCII – входной алфавит.
2. Спецификация – описание формы произвольного слова (восьмеричной константы).  
   *w* *∈ L*, *w = pxs*, где *p* – префикс является цифрой 0; *x* – произвольное слово, составленное из цифр в диапазоне от 0 до 7; *s* – суффикс является либо пустым словом либо словом, состоящим из одного символа либо словом ровно из двух символов подмножества {*l, L, u, U*}. Символы *l* и *L*, а также *u* и *U* – семантически равнозначны (взаимозаменяемы, не изменяют значение константы). Символы *l* и *L* представляют «длинную» константу. Символы *u* и *U* представляют «беззнаковую» константу. Двухсимвольный суффикс *s* допускает только комбинации «длинная» «беззнаковая» либо «беззнаковая» «длинная».
3. [ulUL0-7] – допустимые символы,  
   [^ulUL0-7] – недопустимые символы
4. Регулярное выражение:  
   е = 0[0-7]\*(([uUλ][lLλ])|([lL][uU]))
5. Разбиение класса допустимых символов и всего входного алфавита:  
   [0], [1-7], [uU], [lL], [^ulUL0-7].  
   Символы 0, 7, U, L, ? – представители соответствующих классов, составляют абстрактный алфавит.
6. Система регулярных выражений, определяющих язык:  
   *е* = 0(0|7)\*((λ|U)(λ|L))|(LU);  
   0 = [0];

7 = [1-7];

U = [uU];

L = [lL];

? = [^ulUL0-7].

1. Построить λ- диаграмму (графическое представление НКА) по заданному  
   главному регулярному выражению *е* в абстрактном алфавите {0, 7, U, L, ?}.

0

λ- диаграмма (графическое представление НКА) для *е* = 0(0|7)\*((λ|U)(λ|L))|(LU) в абстрактном алфавите {0, 7, U, L, ?}.

λ

λ

0|7

7

U

L

0|7

λ|U

λ|L

L

U

0|7

4

λ

λ

0

0

1

2

3

5

6

λ|U

λ|L

L

U

1. Построить ДКА = (Q, Σ, *g*, q0, F).  
   Q – множество состояний;  
   Σ = {0, 7, U, L, ?} – абстрактный алфавит;  
   q0=[{0}]=[0] – начальное состояние, q0 ∈Q;  
   F – множество финальных (или заключительных) состояний, F ⊆ Q;  
   *g*: Q×Σ →Q – функция перехода ДКА,   
   *g*(q, a)=[*μ* (q, a)], где *μ*: Q×Σ →2X – функция перемещения ДКА, X = {0,1,2,3, 4, 5, 6} = [0-6] – множество состояний λ- диаграммы (НКА).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| База Данных: q' = [*μ* (q, a)] | q | 0 | 7 | U | L | ? |
| q0 = [0] = {0} | q0 | q1 |  |  |  |  |
| q1 = [1] = {1, 2, 3, 4, 6} | q1 | q2 | q2 | q3 | q4 |  |
| q2 = [2] = {2, 3, 4, 6} | q2 | q2 | q2 | q3 | q4 |  |
| q3 = [4] = {4,6} | q3 |  |  |  | q5 |  |
| q4 = [5] = {5,6} | q4 |  |  | q5 |  |  |
| q5 = [6] = {6} | q5 |  |  |  |  |  |
| q6 = [{}] = {} | q6 |  |  |  |  |  |

q0 = [0] = {0} = 0

*g*(q0, 0) = [*μ* (q0, 0)] ) = [*μ* ({0}, 0)] ) = [*μ* (0, 0)] = [{1}] = [1] = {1, 2, 3, 4, 6} = q1

*g*(q0, 7) = [*μ* (q0, 7)] ) = [*μ* ({0}, 7)] ) = [*μ* (0, 7)] = [{}] = { ø }

*g*(q0, U) = [*μ* (q0, U)] ) = [*μ* ({0}, U)] ) = [*μ* (0, U)] = [{}] = { ø }

*g*(q0, L) = [*μ* (q0, L)] ) = [*μ* ({0}, L)] ) = [*μ* (0, L)] = [{}] = { ø }

*g*(q0, ?) = [*μ* (q0, ?)] ) = [*μ* ({0}, ?)] ) = [*μ* (0, ?)] = [{}] = { ø }

*g*(q1, 0) = [*μ* (q1, 0)] ) = [*μ* ({1, 2, 3, 4, 6}, 0)] ) = [{2}] = [2] = {2, 3, 4, 6} = q2

*g*(q1, 7) = [*μ* (q1, 7)] ) = [*μ* ({1, 2, 3, 4, 6}, 7)] ) = [{2}] = [2] ={2, 3, 4, 6} =q2

*g*(q1, U) = [*μ* (q1, U)] ) = [*μ* ({1, 2, 3, 4, 6}, U)] ) = [{4}] = [4] = {4, 6} = q3

*g*(q1, L) = [*μ* (q1, L)] ) = [*μ* ({1, 2, 3, 4, 6}, L)] ) = [{5, 6}] = [5, 6] = {5, 6} = q4

*g*(q1, ?) = [*μ* (q1 ?)] ) = [*μ* ({1, 2, 3, 4, 6}, ?)] ) = [{}] = { ø }

*g*(q2, 0) = [*μ* (q2, 0)] ) = [*μ* ({2, 3, 4, 6}, 0)] ) = [{2}] = [2] = {2, 3, 4, 6} = q2

*g*(q2, 7) = [*μ* (q2, 7)] ) = [*μ* ({2, 3, 4, 6}, 7)] ) = [{2}] = [2] = {2, 3, 4, 6} = q2

*g*(q2, U) = [*μ* (q2, U)] ) = [*μ* ({2, 3, 4, 6}, U)] ) = [{4}] = [4] = {4, 6} = q3

*g*(q2, L) = [*μ* (q2, L)] ) = [*μ* ({2, 3, 4, 6}, L)] ) = [{5, 6}] = [5, 6] = {5, 6} = q4

*g*(q2, ?) = [*μ* (q2 ?)] ) = [*μ* ({2, 3, 4, 6}, ?)] ) = [{}] = { ø }

*g*(q3, 0) = [*μ* (q3, 0)] ) = [*μ* ({4, 6}, 0)] ) = [{}] = { ø }

*g*(q3, 7) = [*μ* (q3, 7)] ) = [*μ* ({4, 6}, 7)] ) = [{}] = { ø }

*g*(q3, U) = [*μ* (q3, U)] ) = [*μ* ({4, 6}, U)] ) = [{}] = { ø }

*g*(q3, L) = [*μ* (q3, L)] ) = [*μ* ({4, 6}, L)] ) = [{6}] = [6] = {6} = q5

*g*(q3, ?) = [*μ* (q3, ?)] ) = [*μ* ({4, 6}, ?)] ) = [{}] = { ø }

*g*(q4, 0) = [*μ* (q4, 0)] ) = [*μ* ({5, 6}, 0)] ) = [{}] = { ø }

*g*(q4, 7) = [*μ* (q4, 7)] ) = [*μ* ({5, 6}, 7)] ) = [{}] = { ø }

*g*(q4, U) = [*μ* (q4, U)] ) = [*μ* ({5, 6}, U)] ) = [{6}] = [6] = {6} = q5

*g*(q4, L) = [*μ* (q4, L)] ) = [*μ* ({5, 6}, L)] ) = [{}] = { ø }

*g*(q4, ?) = [*μ* (q4, ?)] ) = [*μ* ({5, 6}, ?)] ) = [{}] = { ø }

*g*(q5, 0) = [*μ* (q5, 0)] ) = [*μ* ({6}, 0)] ) = [{}] = { ø }

*g*(q5, 7) = [*μ* (q5, 7)] ) = [*μ* ({6}, 7)] ) = [{}] = { ø }

*g*(q5, U) = [*μ* (q5, U)] ) = [*μ* ({6}, U)] ) = [{}] = { ø }

*g*(q5, L) = [*μ* (q5, L)] ) = [*μ* ({6}, L)] ) = [{}] = { ø }

*g*(q5, ?) = [*μ* (q5, ?)] ) = [*μ* ({6}, ?)] ) = [{}] = { ø }

Q={q0,q1,q2,q3, q4,q5,q6} =[0-6]

F = {q| q ∈Q, n-1∈X, n=|X|, n-1 ∈q}.

n-1 – финальное состояние λ- диаграммы (НКА).

F={q0,q1,q2,q3, q4,q5}. Любое q ∈F называется состоянием «допуска».

Cостояние q6={} называется состоянием «ошибки».

ДКА называется распознавателем или акцептором лексем-идентификаторов.

0|7

4

0|7

0

0

1

2

3

5

U

L

L

U

6

Соглашения по обозначениям.

Пустые клетки в автоматной таблице представляют состояние «ошибки» q6.

Так как q6 = {} – пустое множество, то оно является выделенной константой в семействе подмножеств 2X, где X – множество состояний λ- диаграммы (НКА). Поэтому пустое множество в Базе Данных всегда последнее и его идентификация (назначение индекса) откладывается до окончания итерационного процесса.

Количество «пустых клеток» обычно не меньше, чем «не пустых», отражающих текущие допустимые состояния ДКА. Чтобы не загромождать диаграмму ДКА,

дуги (q, x, q') и петли (q', x, q'), где q' – состояние «ошибки», q – допустимое состояние, не рисуются, но подразумеваются.