УДК 159.854

Студ. С.И. Тумаш
Науч. рук. асс. В.С. Хворост
(кафедра информатики и веб-дизайна БГТУ)

РЕГУЛЯРНЫЕ ВЫРАЖЕНИЯ КАК КОНЕЧНЫЕ АВТОМАТЫ

Регулярные выражения это актуальный инструмент, нашедший применение во многих языках программирования. Их классическая реализация основана на конечных автоматах, где на первом этапе выражение преобразуется в эквивалентный КА. Рассмотрим часть этого алгоритма, где выражение преобразуется в эквивалентный недетерминированный конечный автомат (НКА). Как утверждает первая часть теоремы Клини, для каждого регулярного выражения существует эквивалентный, то есть принимающий тот же язык (множество символьных последовательностей), конечный автомат [1]. Мы будем использовать наиболее интуитивно понятное представление конечного автомата – граф переходов, где состояниям соответствуют вершины, а переходы представлены в виде ребер.

Автомат, эквивалентный выражению, состоящему из некого символа, имеет две вершины-состояния: s0 (начальное) и sF (конечное), а также дугу-переход от s0 к sF с соответствующим символом. Рассмотрим на примере выражения d.

Пусть A и B – конечные автоматы, эквивалентные выражениям R1 и R2 соответственно. Начальные состояния у них a0 и b0, конечные – aF, bF. Эпсилон-переходом называется переход, который может быть выполнен автоматом самопроизвольно, без входного символа. Тогда основные операторы, используемые в регулярных выражениях, определяются для эквивалентных автоматов следующим образом:

– объединение: Автомат C, эквивалентный R1R2, включает A и B, с начальным состоянием a0, эпсилон-переходом из aF в b0 (рисунок 1).

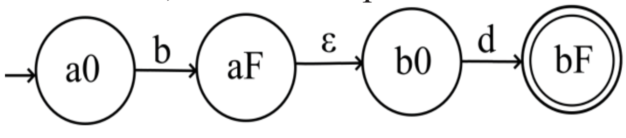

Рисунок 1 – Автомат для выражения bd

– конкатенация: Автомат C, эквивалентный R1|R2, включает A и B, с новым начальным состоянием c0, новым конечным состоянием cF и эпсилон-переходами от c0 к a0 и b0, а также от aF и bF к cF (рисунок 2).

– замыкание Клини: Автомат C, эквивалентный R1*, включает A с новым начальным состоянием c0 и новым конечным состоянием cF, эпсилон-переходами от c0 к a0 и cF, и от aF к a0, а также от aF к cF (рисунок 2).

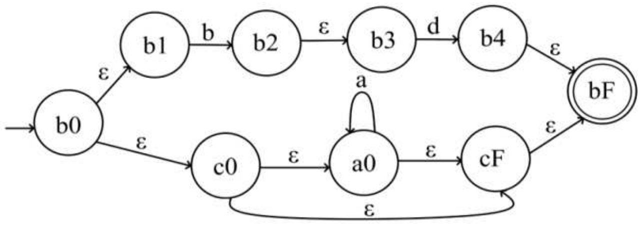

Рисунок 2 – Автомат для выражения (bd)|(a*)

Займемся удалением эпсилон-переходов. Для начала все состояния, в которые существуют только эпсилон-переходы, удаляются (рисунок 3).

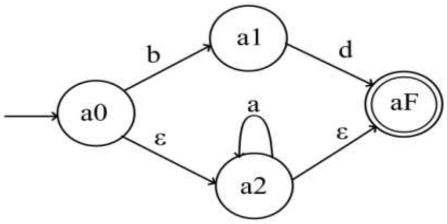

Рисунок 3 – Демонстрация на примере автомата с рисунка 2

Для более сложных случаев применяется другой подход. Пусть a и b – узлы, между которыми существует эпсилон-переход (рисунок 4). Необходимо:

- найти все переходы из b;
- продублировать их, добавив аналогичные переходы из a;
- удалить эпсилон-переход;
- обозначить состояние b как начальное, если таковым является a;
- обозначить состояние a как конечное, если таковым является b;

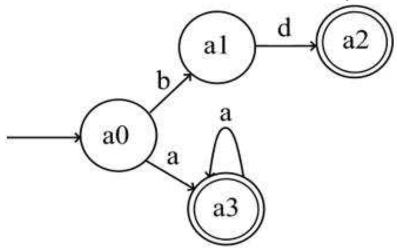

Рисунок 4 – Демонстрация на примере автомата с рисунка 3

В ходе проделанной работы мы изучили алгоритм преобразования регулярного выражения в НКА, этого достаточно для реализации регулярных выражений. На практике обычно добавляют еще один шаг с преобразованием НКА в детерминированный автомат, который более прост для программной реализации, но в данной работе этот шаг избыточен.

ЛИТЕРАТУРА

1. М. Ниват. Regular expressions into finite automata //Theoretical Computer Science – Is. 120.–Великобритания: Lsevier, 1993.