CS361, Fall 1995

Instructor: Guo-Qiang Zhang

Lecture Notes of 09/29/95

Topic: Regular Expressions and Regular Languages

Scribed by Park and Pecqueur

1 Regular Expressions

Regular expressions are a way to describe regular languages. The notation involves a combination of strings of symbols from some alphabet Σ, parentheses, and the operators $+$, $\cdot$, and *. Formally we can describe a **regular expression** as follows: Let Σ be a given alphabet. **1.** $\emptyset$, λ and a ($\in \Sigma$) are all regular expressions. These are called **primitive regular expressions**. **2.** If r_1 and r_2 are regular expressions, so are $r_1 + r_2$, $r_1 \cdot r_2$, r_1^*, and (r_1). **3.** A string is a regular expression if and only if it can be derived from the primitive regular expressions by a finite number of applications of the rules in (2).

Note: We don't have Conjunction and Complement for regular expressions, but DFA's can handel it.

2 Designing Regular Expressions

Let $L_1 = \{a^{2n}b^{2m+1} \mid n \geq 0, m \geq 0\}$. Then a regular expression for L_1 $(aa)^*b(bb)^*$.

Let $L_2 = \{w \mid w \in 0, 1^*, w$ has at least one pair of consecutive $0's\}$. Then a regular expression for L_2 is

$$(0+1)^*00(0+1)^*.$$

Let $L_3 = \{0^i1^j \mid$ exactly one of i, j is odd$\}$. Then a regular expression for L_3 is

$$(00)^*(0+1)(11)^*.$$

The following table demonstrates the differences between regular expressions:

	0^*1^*	$(01)^*$	$0^* + 1^*$	$(0 + 1)^*$
0^*1^*	X	001	01	010
$(01)^*$	0101	X	0101	0110
$(0^* + 1^*)$	01	01	X	01
$(0 + 1)^*$	010	0	01	X

Where the entry in a particular box corresponds to a string which is in one of the regular expression, but not the other.

3 Regular Expressions from DFA's

If you have a DFA you can always find a regular expression describing the language accepted by it.

Figure 1. $(a + b)^*$ is the language accepted by

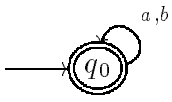


Figure 2. $(b + aa^*b)^*$ is the language accepted by

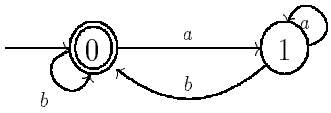


Figure 3. $b^*a(a + bb^*a)^*$ is the language accepted by

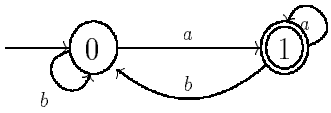


Figure 4. $(a + b)(a + b)^*$ describes the language accepted by

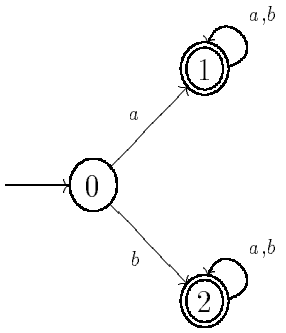
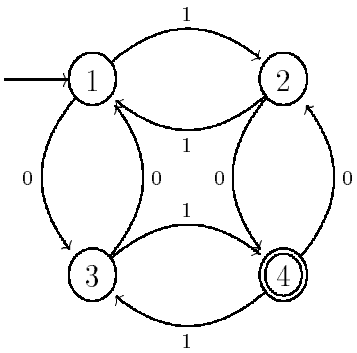


Figure 5. What is a regular expression which describes the language accepted by the following DFA?



4 Arden's Lemma

Arden's Lemma is a method to convert complex DFA's to regular expressions. The lemma states that

> An equation of the form:
>
> $$X = XA + B$$
>
> with $\lambda \notin A$ has the unique solution $X = BA^*$.

Here is an example of Arden's Lemma to convert the DFA given in Figure 5 to a regular expression.

$$X_1 = X_2 1 + X_3 0 + \lambda$$
$$X_2 = X_1 1 + X_4 0$$
$$X_3 = X_1 0 + X_4 1$$
$$X_4 = X_2 0 + X_3 1$$

So we have

$$\begin{aligned} X_4 &= (X_1 1 + X_4 0)0 + (X_1 0 + X_4 1)1 \\ &= X_1 10 + X_4 00 + X_1 01 + X_4 11 \\ &= X_4(00 + 11) + X_1(10 + 01) \end{aligned}$$

Applying the lemma yields:

$$X_4 = X_1(10 + 01)(00 + 11)^*$$

Similarly, for X_1, we have

$$\begin{aligned} X_1 &= X_2 1 + X_3 0 + \lambda \\ &= (X_1 1 + X_4 0)1 + (X_1 0 + X_4 1)0 + \lambda \\ &= X_1(11 + 00) + X_4(01 + 10) + \lambda \end{aligned}$$

Applying the lemma again yields

$$X_1 = [X_4(01 + 10) + \lambda](11 + 00)^*.$$

Now substitute this back to the earlier solution for X_4, we have

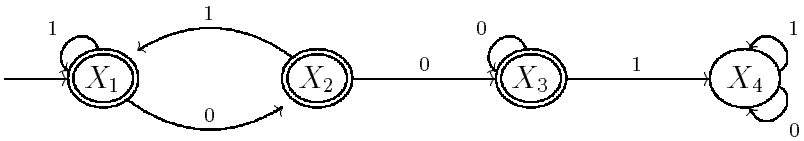
$$\begin{aligned} X_4 &= [X_4(01 + 10) + \lambda](11 + 00)^*(10 + 01)(00 + 11)^* \\ &= X_4(01 + 10)(11 + 00)^*(10 + 01)(00 + 11)^* + (11 + 00)^*(10 + 01)(00 + 11)^* \end{aligned}$$

and so the final solution is

$$X_4 = (11 + 00)^*(01 + 10)(00 + 11)^*[(10 + 01)(11 + 00)^*(01 + 10)(00 + 11)^*]^*$$

For the DFA pictured in page 45 Fig. 2.5 (see below), we can derive a regular expression in a similar way, which is

$$(1 + 01)^* + 1^*0(11^*0)^* + 1^*0(11^*0)^*00^*.$$


1
1
0
X_1
X_2
X_3
X_4
0
1
0
1
0