CS361, Fall 1995

Instructor: Guo-Qiang Zhang

Lecture Notes of 09/22/95

Topic: Strings and Languages

Scribed by Chandra and Gonsalves

1 Introduction

At the basis of Automata theory is the notion of strings and formal languages. Strings are also sometimes called words. Formal Languages have to be specified in very precise way so that they can be accepted by computers. Therefore Formal languages are studied very rigorously using set notation. Basically, Formal Languages are used to measure the power of certain categories of machines. These concepts are also used in the study of compilers.

Σ is a **finite, non empty set** of symbols.

e.g., $\Sigma = \{0, 1, 2, \ldots, 9\}$

We use symbols in Σ to make up strings. A string over Σ is a sequence of symbols from Σ.

e.g., $\Sigma = \{ \text{0,1} \}$

e.g., of strings from the above alphabet, $000, 10, 11$ and 101010.

A Language L is a **set of strings** over Σ.

2 Operations on Strings

2.1 Concatenation

The **concatenation** of two strings u and v is the string obtained by appending the symbols of v to the end of u. For example, if $u = 010$ and $v = 11$, then $uv = 01011$.

Concatenation is not commutative: $uv \neq vu$.

2.2 Reverse

The **reverse** of a string is obtained by writing the symbols in reverse order.

e.g., u = 1101, $u^R = 1011$

2.3 The empty string

The empty string is a string with no symbols at all. It is denoted by λ, and is characterized by the following two properties: $\lambda w = w$ and $w\lambda = w$ for any string w.

2.4 Prefix and Suffix

Prefix is the initial portion of the word and **Suffix** is the later portion of the word.

If w = uv, then u is called a prefix of w and v is called a suffix of w.

e.g., List of all prefixes of the string 1101

1,11,110, 1101, λ

e.g., List of all suffixes of the string 1101

1, 01, 101, 1101, λ

2.5 Length of a string

The length of a string w is denoted by $\mid w \mid$.

The length of the empty string λ , $\mid \lambda \mid = 0$.

e.g.,

$\mid abaa \mid = 4$

$\mid uw \mid = \mid u \mid + \mid w \mid$

$\mid u^R \mid = \mid u \mid$

3 Operations on Languages

A Language is a set of strings. A Language is defined very generally as a subset of Σ^*.

Since Languages are sets, the union, intersection and difference of two languages are defined.

The Union of two Languages L_1, L_2 is denoted symbolically as, $L_1 \cup L_2$. The Union of two Languages contains all the strings belonging to L_1 and L_2.

The intersection of two Languages L_1, L_2 is defined symbolically as, $L_1 \cap L_2$. The Intersection of two languages contains only the strings belonging to both L_1 and L_2.

The difference of two Languages L_1, L_2 is defined symbolically as, $L_1 - L_2$. The Difference of two languages contains the strings that belong to L_1 but not to L_2.

The reverse L^R of a language L is defined as $L^R = \{w^R \mid w \in L\}$

The concatenation L_1L_2 is defined as $L_1L_2 = \{$ uw | u $\in L_1$ & w $\in L_2\}$

Σ^* contains all the strings from the alphabet Σ

$\Sigma^* = \{$ w | w is a string over Σ $\}$

L is a language over Σ iff L $\subseteq \Sigma^*$

The * operator can be applied to any language L.

L^* is defined as the collection of all strings formed by the concatenation of zero or more strings from L.

e.g.,

$\lambda = w^0$

$w = w^1$

$ww = w^2$

$www = w^3$

From the above examples we can see that , $\lambda \in L^*$ for any L.

For e.g., if $L = \{a^2, a^3\}$, $L^* = \{a\}^* - \{a\}$

This can be proved by showing the set equalities in both directions.

$(L^*)* = L^*$

4 Review Questions

Question	Answer
$\emptyset^* = \emptyset$	FALSE (because $\lambda \in \emptyset^*$)
$\emptyset^* = \{\lambda\ \}$	TRUE (because $\{\lambda\}^*$ contains λ
$\{\lambda\}^* = \emptyset$	FALSE
$\{\lambda\}^* = \{\lambda\}$	TRUE
$\{\lambda\} = \emptyset$	FALSE
$\{\lambda\} = \{\emptyset\}$	FALSE (because $\{\emptyset\}$ is not a language)