THEORY OF AUTOMATA
&
FORMAL LANGUAGES

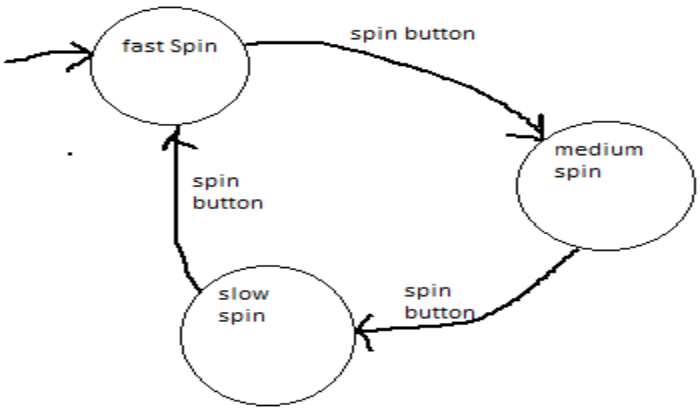
HANDOUTS 01

Department of Information Technology
Hazara University, Mansehra

What does automata mean?

It is the plural of automaton, and it means "something that works automatically".

"Finite" refer to the finite number of" state(s)" (a condition at particular time) in the automata. For e.g spinner of the washing machine(consider it as automata) can be in any three states(represented by circle) at particular time. i.e. fast spin, medium spin, slow spin. A transition(represented by arrow) from one state to another can be triggered through a spin button. This can be represented as following.



Introduction to languages

There are two types of languages

- Formal Languages (Syntactic languages)
- Informal Languages (Semantic languages)

Alphabets

Definition

A finite non-empty set of symbols (called letters), is called an alphabet. It is denoted by Σ (Greek letter sigma).

Example

Σ = {a,b}

Σ = {0,1} (*important as this is the language which the computer understands*.)

Σ = {i,j,k}

Note Certain version of language ALGOL has 113 letters.

Σ (alphabet) includes letters, digits and a variety of operators including sequential operators such as GOTO and IF

Strings

Definition

Concatenation of finite number of letters from the alphabet is called a string.

Example

If Σ = {a,b} then

a, abab, aaabb, abababababababab

Note

Empty string or null string: Sometimes a string with no symbol at all is used, denoted by (Small Greek letter Lambda) λ or (Capital Greek letter Lambda) Λ, is called an empty string or null string.

The capital lambda will mostly be used to denote the empty string, in further discussion.

Words

Definition

Words are strings belonging to some language.

Example

If Σ = {x} then a language L can be defined as

L={x^n : n=1,2,3,⋯..} or L={x,xx,xxx,⋯.}

Here x,xx,⋯ are the words of L

Note: All words are strings, but not all strings are words.

Valid/In-valid alphabets

While defining an alphabet, an alphabet may contain letters consisting of group of symbols for example Σ1= {B,aB, bab, d}.

Now consider an alphabet

Σ2= {B, Ba, bab, d} and a string BababB.

This string can be tokenized in two different ways

(Ba), (bab), (B)

(B), (abab), (B)

Which shows that the second group cannot be identified as a string, defined over

Σ = {a, b}.

As when this string is scanned by the compiler (Lexical Analyzer), first symbol B is identified as a letter

belonging to Σ, while for the second letter the lexical analyzer would not be able to identify, so while defining an alphabet it should be kept in mind that ambiguity should not be created.

Remarks

While defining an alphabet of letters consisting of more than one symbols, no letter should be started with the letter of the same alphabet *i.e.* one letter should not be the prefix of another. However, a letter may be ended in a letter of same alphabet.

Conclusion

Σ1= {B, aB, bab, d}

Σ2= {B, Ba, bab, d}

Σ1 is a valid alphabet while Σ2 is an in-valid alphabet.

Length of Strings

Definition

The length of string s, denoted by |s|, is the number of letters in the string.

Example

Σ={a,b}

s=ababa

|s|=5

Example

Σ = {B, aB, bab, d}

s=BaBbabBd

Tokenizing=(B), (aB), (bab), (B), (d)

|s|=5

Reverse of a String

Definition

The reverse of a string s denoted by Rev(s) or sr, is obtained by writing the letters of s in reverse order.

Example

If s=abc is a string defined over Σ={a,b,c}

then Rev(s) or sr = cba

Example

Σ = {B, aB, bab, d}

s=BaBbabBd

Rev(s)=dBbabaBB