Department of Computer & Information Sciences
Pakistan Institute of Engineering and Applied Sciences

Concepts of Automata Theory

Umar Faiz
http://www.pieas.edu.pk/umarfaiz/cis317

Concepts of Automata Theory

Three basic concepts

Alphabet	a set of symbols
Strings	a list of symbols from an alphabet
Language	a set of strings from the same alphabet

Concepts of Automata Theory

Alphabets:

A common way to talk about words, number, pairs of words, etc. is by representing them as strings. To define strings, we start with an alphabet

Concepts of Automata Theory

Alphabets:

Definition: An alphabet is a finite, nonempty set of symbols.
Conventional notation, S.

The term "symbol" is usually not defined.

Example:

$\Sigma_1 = \{a, b, c, d, \ldots, z\}$: the set of letters in English
$\Sigma_2 = \{0, 1, \ldots, 9\}$: the set of (base 10) digits
$\Sigma_3 = \{a, b, \ldots, z, \#\}$: the set of letters plus the special symbol #
$\Sigma_4 = \{(,)\}$: the set of open and closed brackets

Concepts of Automata Theory

Strings:

Definition: A string (or word) is a finite sequence of symbols from some alphabet.

Example - 1011 is a string from alphabet $\Sigma = \{0, 1\}$

- Empty string ε - a string with zero occurrences of symbols
- Length $|w|$ of string w - the number of positions for symbols in w

Examples - |0111|=4, |e|=0…

Set of all strings over S --- denoted as S*. It is not difficult to know that S* = S0 ∪ S1 ∪ S2 ∪ …

Concepts of Automata Theory

Strings:

Example: A string (or word) is a finite sequence of symbols from some alphabet.

abfbz is a string over $\Sigma_1 = \{a, b, c, d, \ldots, z\}$
9021 is a string over $\Sigma_2 = \{0, 1, \ldots, 9\}$
ab#bc is a string over $\Sigma_3 = \{a, b, \ldots, z, \#\}$
))()(() is a string over $\Sigma_4 = \{(,)\}$

Concepts of Automata Theory

Languages:

Definition: A language is a set of strings all chosen from some S*.

If Σ is an alphabet, and $L \subseteq \Sigma^*$, then L is a language over Σ.

Examples:

- The set of all legal English words is a language. Why? What is the alphabet here? Answer: the set of all letters
- A legal program of C is a language. Why? What is the alphabet here? Answer: a subset of the ASCII characters.

Languages

Languages can be used to describe problems with "yes/no" answers, for example:

L1 = The set of all strings over S1 that contain the substring "fool"

L2 = The set of all strings over S2 that are divisible by 7
= {7, 14, 21, …}

L3 = The set of all strings of the form s#s where s is any string over {a, b, …, z}

L4 = The set of all strings over S4 where every (can be matched with a subsequent)

Concepts of Automata Theory

Languages:

Examples:

The set of all strings of n 0's followed by n 1's for $n \geq 0$:

$\{\varepsilon, 01, 0011, 000111, \ldots\}$

- Σ^* is an infinite language for any alphabet.
- ϕ = the empty language (not the empty string e) is a language over any alphabet.
- $\{\varepsilon\}$ is a language over any alphabet (consisting of only one string, the empty string e).

Concepts of Automata Theory

Languages:

Ways to describe languages:

Description by exhaustive listing

- $L1 = \{a, ab, abc\}$ (finite language; listed one by one)
- $L2 = \{a, ab, abb, abbb, \ldots\}$ (infinite language; listed partially)
- $L3 = L(ab^*)$ (infinite language; expressed by a regular expression)

Concepts of Automata Theory

Languages:

Ways to describe languages:

Description by generic elements

L4 = {x | x is over V = {a, b}, begins with a, followed by any number of b, possible none}

(note: $L4 = L3 = L2$)

Description by integer parameters

L5 = {abn | n ≥ 0}

(note: L5=L4=L3=L2)

Concepts of Automata Theory

Operations on Languages:

Languages are sets, and operations of sets may be applied to them:

(1) union - $A \cup B = \{a \mid a \in A \text{ or } a \in B\}$

(2) intersection - $A \cap B = \{a \mid a \in A \text{ and } a \in B\}$

(3) difference - $A - B = \{a \mid a \in A \text{ and } a \notin B\}$

(4) product - $A \times B = \{(a, b) \mid a \in A \text{ and } b \in B\}$

(5) complement - $\bar{A} = \{a \mid a \in U \text{ and } a \notin A\}$

(6) power set - $2^A = \{B \mid B \subset A\}$

U is the universal set, like an alphabet for sets of symbols

Concepts of Automata Theory

Operations on Languages:

The usual set operations

$$\{a, ab, aaaa\} \cup \{bb, ab\} = \{a, ab, bb, aaaa\}$$

$$\{a, ab, aaaa\} \cap \{bb, ab\} = \{ab\}$$

$$\{a, ab, aaaa\} - \{bb, ab\} = \{a, aaaa\}$$

Complement: $\overline{L} = \Sigma^* - L$

$$\overline{\{a, ba\}} = \{\lambda, b, aa, ab, bb, aaa, \ldots\}$$

Concepts of Automata Theory

Operations on Languages:

Reverse:

Definition: $L^R = \{w^R : w \in L\}$

Examples: $\{ab, aab, baba\}^R = \{ba, baa, abab\}$

$$L = \{a^n b^n : n \geq 0\}$$

$$L^R = \{b^n a^n : n \geq 0\}$$

Concepts of Automata Theory

Operations on Languages:

Concatenation:

Definition: $L_1 L_2 = \{xy : x \in L_1, y \in L_2\}$

Example: $\{a, ab, ba\}\{b, aa\}$

$$= \{ab, aaa, abb, abaa, bab, baaa\}$$

Concepts of Automata Theory

Operations on Languages:

Star-Closure (Kleene *):

Definition: $L^* = L^0 \cup L^1 \cup L^2 \cdots$

Example:

$$\{a, bb\}^* = \left\{ \begin{array}{l} \lambda, \\ a, bb, \\ aa, abb, bba, bbbb, \\ aaa, aabb, abba, abbbb, \ldots \end{array} \right\}$$

Concepts of Automata Theory

Operations on Languages:

Positive Closure:

Definition: $L^+ = L^1 \cup L^2 \cup \cdots$

$$= L^* - \{\lambda\}$$

$$\{a, bb\}^+ = \left\{ \begin{array}{l} a, bb, \\ aa, abb, bba, bbbb, \\ aaa, aabb, abba, abbbb, \ldots \end{array} \right\}$$