

CS402 - Theory of Automata Glossary By www.virtualians.pk

Acyclic Graph : A directed graph is said to be acyclic if it contains no cycles.

Algorithm : A detailed and unambiguous sequence of instructions that describes how a computation is to proceed and can be implemented as a program.

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

Automata : The plural of automaton. That's right. The plural of automaton is automata, not automatons. It's a Greek thing. Deal with it.

Automaton : A device that functions automatically. In the context of computer science, it is a model of computation. Input is provided by the user, which is then processed automatically by the automaton to obtain a result. In this sense, an automaton is very much like an algorithm.

Bit : The smallest unit of information; the answer to a yes/no question; the outcome of a coin toss; a 0 or a 1.

Cardinality of the set : The number of elements in a set A , denoted $|A|$, is called the cardinality of the set. A set with no elements (i.e., cardinality equals 0) is called the empty set and is denoted by \emptyset .

CFG : CFG is a collection of the followings 1. An alphabet of letters called terminals from which the strings are formed, that will be the words of the language. 2. A set of symbols called non-terminals, one of which is S , stands for "start here". 3. A finite set of productions of the form non-terminal \rightarrow finite string of terminals and/or non-terminals.

Chomsky Normal Form : If a CFG has only productions of the form

nonterminal \rightarrow string of two nonterminals

or

nonterminal \rightarrow one terminal

then the CFG is said to be in Chomsky Normal Form (CNF).

Chomsky Normal Form (CNF) : If a CFG has only productions of the form nonterminal \rightarrow string of two nonterminals or nonterminal \rightarrow one terminal then the CFG is said to be in Chomsky Normal Form (CNF).

Complement : A set composed of all elements that are not members of another set.

Complement of a language : Let L be a language defined over an alphabet Σ , then the language of strings, defined over Σ , not belonging to L , is called Complement of the language L , denoted by L_c or L' .

Complex Number : A number that has a real component and an imaginary component and is characterized as a point on a plane (instead of the real number line).

Computational Complexity : Computational complexity measures the time and memory resources that a computer requires in order to solve a problem. A somewhat more robust measure may be defined by invoking the Universal Turing Machine. The Computational Complexity of a problem is then defined as the time it takes for the fastest program running on a universal computer (as measured in number of computing steps) to compute the solution to the problem.

Countability : A set A is said to be countable if there exists an onto function f from the set of natural numbers to A . The set is said to be countably infinite if there exists a one-to-one onto function f from the set of natural numbers to A . A set that is not countable is said to be uncountable.

Dead production : A production of the form nonterminal \rightarrow terminal is called a dead production.

Decidability Problems : Following are the decidable problems w.r.t. CFG 1. Whether or not the given CFG generates any word? Problem of emptiness of CFL. 2. Whether or not the given CFG generates the finite language? Problem of finiteness. 3. Whether or not the given string w can be generated by the given CFG? Problem of membership.

Deterministic automata : The output is uniquely determined by the input sequences; that is, you get a definite output given any input. The behavior of such automata can be accurately

predicted if the transfer operator is known and given in the form of a table of a logical function, and if you also know the initial state and the input sequence.

Directed Graphs : A directed graph G is a pair (V, E) , where V is a finite set and E is a relation on V . The elements of V are called nodes or vertices. The elements of E are called edges or arcs.

Discrete : Taking only non-continuous values, e.g., Boolean or natural numbers.

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.

Finite Automaton with output : The machines which generates an output string corresponding to each input string. Such machines are called machines with output.

Finite State Machine : (FSM or "Finite State Automaton", "transducer") An abstract machine consisting of a set of states (including the initial state), a set of input events, a set of output events, and a state transition function. The function takes the current state and an input event and returns the new set of output events and the next state. Some states may be designated as "terminal states". The state machine can also be viewed as a function which maps an ordered sequence of input events into a corresponding sequence of (sets of) output events.

Finite-State Automaton (FSA) : The simplest computing device. Although it is not nearly powerful enough to perform universal computation, it can recognize regular expressions. FSAs are defined by a state transition table that specifies how the FSA moves from one state to another when presented with a particular input. FSAs can be drawn as graphs.

Formal language : A language that can be defined by a formal system, that is, by a system that has a finite number of axioms and a finite number of inference rules, is said to be a formal language.

Functions : A function f from A to B , denoted by $f: A \rightarrow B$, is a relation from A to B , whose range $f(x)$ at each x in A has cardinality 0 or 1. $f(x)$ is said to be defined if it has cardinality 1, that is, if $f(x) = \{y\}$ for some y . In such a case, $f(x)$ is said to have the value of y , written $f(x) = y$. Otherwise, $f(x)$ is said to be undefined.

Graph : A construct that consists of many nodes connected with edges. The edges usually represent a relationship between the objects represented by the nodes. For example, if the nodes are cities, then the edges may have numerical values that correspond to the distances between the cities. A graph can be equivalently represented as a matrix.

head : The first item of a list.

Irrational Number : A real number that cannot be represented as a fraction.

Kleene Star Closure :

Given Σ , then the Kleene Star Closure of the alphabet Σ , denoted by Σ^* , is the collection of all strings defined over Σ , including Λ .

It is to be noted that Kleene Star Closure can be defined over any set of strings.

Examples

If $\Sigma = \{x\}$

Then $\Sigma^* = \{\Lambda, x, xx, xxx, xxxx, \dots\}$

If $\Sigma = \{0,1\}$

Then $\Sigma^* = \{\Lambda, 0, 1, 00, 01, 10, 11, \dots\}$

If $\Sigma = \{aaB, c\}$

Then $\Sigma^* = \{\Lambda, aaB, c, aaBaaB, aaBc, caaB, cc, \dots\}$

Kleene's Theorem :

If a language can be expressed by

1. FA or
2. TG or

3. RE

then it can also be expressed by other two as well.

It may be noted that the theorem is proved, proving the following three parts

Kleene's Theorem Part I

If a language can be accepted by an FA then it can be accepted by a TG as well.

Kleene's Theorem Part II

If a language can be accepted by a TG then it can be expressed by an RE as well.

Kleene's Theorem Part III

If a language can be expressed by a RE then it can be accepted by an FA

Leaf :

A node in a graph with no successors is called a leaf.

Left most derivation :

The derivation of a word w , generated by a CFG, such that at each step, a production is applied to the left most nonterminal in the working string, is said to be left most derivation.

It is to be noted that the nonterminal that occurs first from the left in the working string, is said to be left most nonterminal.

Live production :

A production of the form

nonterminal \rightarrow string of two nonterminals

is called a live production.

Mealy machine :

A Mealy machine consists of the following

1. A finite set of states q_0, q_1, q_2, \dots where q_0 is the initial state.
2. An alphabet of letters $S = \{a, b, c, \dots\}$ from which the input strings are formed.
3. An alphabet $G = \{x, y, z, \dots\}$ of output characters from which output strings are generated.
4. A pictorial representation with states and directed edges labeled by an input letter along with an output character. The directed edges also show how to go from one state to another corresponding to every possible input letter.

(It is not possible to give transition table in this case.)

Note

It is to be noted that since, similar to Moore machine, in Mealy machine no state is designated to be a final state, so there is no question of accepting any language by Mealy machine. However in some cases the relation between an input string and the corresponding output string may be identified by the Mealy machine. Moreover, the state to be initial is not important as if the machine is used several times and is restarted after some time, the machine will be started from the state where it was left off.

Model of Computation : An idealized version of a computing device that usually has some simplifications such as infinite memory. A Turing machine, the lambda calculus, and Post production systems are all models of computation.

Moore machine :

A Moore machine consists of the following

1. A finite set of states q_0, q_1, q_2, \dots where q_0 is the initial state.
2. An alphabet of letters $S = \{a, b, c, \dots\}$ from which the input strings are formed.
3. An alphabet $G = \{x, y, z, \dots\}$ of output characters from which output strings are generated.

4. A transition table that shows for each state and each input letter what state is entered the next.
5. An output table that shows what character is printed by each state as it is entered.

Note

It is to be noted that since in Moore machine no state is designated to be a final state, so there is no question of accepting any language by Moore machine. However in some cases the relation between an input string and the corresponding output string may be identified by the Moore machine. Moreover, the state to be initial is not important as if the machine is used several times and is restarted after some time, the machine will be started from the state where it was left off. Following are the examples

Myhill Nerode theorem :

For a language L , defined over an alphabet Σ ,

1. L partitions Σ^* into distinct classes.
2. If L is regular then, L generates finite number of classes.
3. If L generates finite number of classes then L is regular.

Natural Number : Any of the standard counting numbers; a positive integer.

Non-Terminals : The symbols that must be replaced by other things are called non-terminals.

Nondeterministic Finite Automaton (NFA) :

An NFA is a TG with a unique start state and a property of having single letter as label of transitions. An NFA is a collection of three things

- 1) Finite many states with one initial and some final states
- 2) Finite set of input letters, say, $S = \{a, b, c\}$

3) Finite set of transitions, showing where to move if a letter is input at certain state (\emptyset is not a valid transition), there may be more than one transition for certain letters and there may not be any transition for certain letters.

Null Production :

The production of the form

nonterminal $\rightarrow L$ is said to be null production.

Nullable Production :

A production is called nullable production if it is of the form

$N \rightarrow L$

or

there is a derivation that starts at N and leads to L i.e. $N_1 \rightarrow N_2, N_2 \rightarrow N_3, N_3 \rightarrow N_4, \dots, N_n \rightarrow L$, where N, N_1, N_2, \dots, N_n are non terminals.

PLUS Operation (+) :

Plus Operation is same as Kleene Star Closure except that it does not generate Λ (null string), automatically.

Example

If $\Sigma = \{0,1\}$

Then $\Sigma^+ = \{0, 1, 00, 01, 10, 11, \dots\}$

If $\Sigma = \{aab, c\}$

Then $\Sigma^+ = \{aab, c, aabaab, aabc, caab, cc, \dots\}$

POP and STACK : POP and STACK POP is an operation that takes out a letter from the top of the STACK. The rest of the letters are moved one location up.

Prime Number : A natural number that can be evenly divided only by itself and 1.

Probability : The likelihood that a random event will occur.

Productions : The grammatical rules are often called productions.

Program : An algorithm that is written in a programming language for execution on a physical computer.

Pseudo theorem :

For three languages P,Q and R, while $PQ=R$ the language Q must be regular if both P and R are regular.

(Note: It is to be noted that since this theorem is not true, so the theorem is called pseudo theorem.)

Pumping Lemma :

Let L be any infinite regular language (that has infinite many words), defined over an alphabet Σ then there exist three strings x, y and z belonging to Σ^* (where y is not the null string) such that all the strings of the form $x^n y^n z^n$ for $n=1,2,3, \dots$ are the words in L.

Pumping Lemma version II :

Let L be an infinite language accepted by a finite automaton with N states, then for all words w in L that have length more than N, there are strings x,y and z (y being non-null string) and $\text{length}(x) + \text{length}(y) \leq N$ s.t.

$w = xyz$ and all strings of the form $x^n y^n z^n$ are in L for $n = 1,2,3, \dots$

PUSH : PUSH A PUSH operator adds a new letter at the top of STACK.

PUSHDOWN AUTOMATON (PDA) :

Pushdown Automaton (PDA), consists of the following

1. An alphabet S of input letters.
2. An input TAPE with infinite many locations in one direction. Initially the input string is placed in it starting from first cell, the remaining part of the TAPE is empty.
3. An alphabet G of STACK characters.
4. A pushdown STACK which is initially empty, with infinite many locations in one direction. Initially the STACK contains blanks.
5. One START state with only one out-edge and no in-edge.
6. Two halt states i.e. ACCEPT and REJECT states, with in-edges and no out-edges.
7. A PUSH state that introduces characters onto the top of the STACK.
8. A POP state that reads the top character of the STACK, (may contain more than one out-edges with same label).
9. A READ state that reads the next unused letter from the TAPE, (may contain more than one out-edges with same label).

PUSHDOWN STACK or PUSHDOWN STORE :

It is a place where the input letters can be placed until these letters are referred again. It can store as many letters as one can in a long column.

Rational Number : A number that can be expressed as a fraction.

Real Number : Any number that can be represented with a potentially infinite decimal expansion to the right of the decimal point. Natural, rational, irrational, and incomputable numbers are all real numbers.

Recursive : Strictly speaking, a set or function is recursive if it is computable; however, in the usual sense of the word, a function is said to be recursive if its definition make reference to itself. For example, factorial can be defined as $x! = x * (x - 1)!$ with the base case of $1!$ equal to 1

Recursive definition of Regular Expressions :

Recursive definition of Regular Expression (RE)

Step 1: Every letter of Σ including Λ is a regular expression.

Step 2: If r_1 and r_2 are regular expressions then

- a. (r_1)
- b. $r_1 r_2$
- c. $r_1 + r_2$ and
- d. r_1^*

are also regular expressions.

Step 3: Nothing else is a regular expression.

Regular Expression : A definition for a class of strings that can be recognized by a finite-state automaton. An example of a class of strings that is regular would be legal mathematical expressions using only "+" and digits. An example that is not regular is the same legal mathematical expressions as before, but with properly nested parentheses.

Regular Languages :

The language generated by any regular expression is called a regular language.

It is to be noted that if r_1, r_2 are regular expressions, corresponding to the languages L_1 and L_2 then the languages generated by $r_1 + r_2, r_1 r_2$ (or $r_2 r_1$) and r_1^* (or r_2^*) are also regular languages.

Rooted Acyclic Graphs : An acyclic graph is said to be rooted if exactly one of its nodes, called the root, has no predecessors.

Set : A collection of things, usually numbers. Sets may be infinite in size.

Sets :

A set is a collection of elements. The order or repetition of the elements is immaterial. Notation of the form $\{ x \mid x \text{ satisfies the property } Q \}$ is used for specifying the set of all elements x that satisfy property Q . Finite sets are also specified by explicitly listing their members between braces.

State : No, not the place in which you live, or some country somewhere in the world. Rather, state is used in the context of an automaton and refers to what is remembered by the finite control of the automaton at any point in a computation. So, "state" refers in some sense to the "state of being" of the automaton at any given point in its computation.

String : Any sequence of letters, numbers, digits, bits, or symbols.

Terminals : The symbols that can't be replaced by anything are called terminals.

Total language tree : For a given CFG, a tree with the start symbol S as its root and whose nodes are working strings of terminals and non-terminals. The descendants of each node are all possible results of applying every production to the working string. This tree is called total language tree.

Transition Graph :

Transition Graph

Definition:

A Transition graph (TG), is a collection of the followings

- 1) Finite number of states, at least one of which is start state and some (maybe none) final states.
- 2) Finite set of input letters (Σ) from which input strings are formed.

Finite set of transitions that show how to go from one state to another based on reading specified substrings of input letters, possibly even the null string

Tree : A rooted, acyclic, directed graph is called a tree if each of its nodes, excluding the root, has exactly one predecessor.

Turing Machine : A model of computation that uses an underlying finite-state automaton but also has an infinite tape to use as memory. Turing machines are capable of universal computation.

Unary alphabet : Unary alphabet $\{1\}$ is a unary alphabet.

Unit production : Unit production The productions of the form nonterminal \rightarrow one nonterminal, is called the unit production. Following is an example showing how to eliminate the unit productions from a given CFG.

Universal Computer : A computer that is capable of universal computation, which means that given a description of any other computer or program and some data, it can perfectly emulate this second computer or program. Strictly speaking, home PCs are not universal computers because they have only a finite amount of memory. However, in practice, this is usually ignored.

Vector : A one-dimensional array of numbers that can be used to represent a point in a multidimensional space.

XOR : The exclusive-or function; given two Boolean inputs, the output of XOR is 1 if and only if the two inputs are different; otherwise, the output is 0.