



# Compiler Construction

## Humanly Project Documentation

BS-CS-5B

**Course Code:** CS-3102

**Instructor:** Ma'am Raazia Sosan

**Student Names:**

Aheed Uddin Basri (cs182010)

Muhammad Babar Siddiqi (cs182006)

Aaisha Motan (cs182009)

## Table of Contents

Introduction .....	3
Lexical Aspects .....	3
Rules for Identifiers.....	3
Constants .....	3
Integer & Decimal Constants .....	3
String Constant.....	3
<b>Escape Sequence</b> .....	4
<b>Meaning</b> .....	4
Operators and Punctuators .....	4
Comments .....	4
Reserved Keywords.....	4
<b>Humanly</b> .....	4
<b>C++</b> .....	4
Data Types.....	5
<b>Humanly</b> .....	5
<b>C++</b> .....	5
Humanly Best Practices/Naming Conventions .....	5
NFA's, DFA's and Regular Expressions .....	6
Data Types.....	6
Reserevd Keywords.....	8
Punctuators.....	18
Identifiers .....	20

# Humanly

## Introduction

Humanly as its name is a language that is designed to be as human-like as possible. It is inspired by the natural English keywords that people use in their daily life for explaining things very often. The purpose of this language is to reduce ambiguity related to different keywords and maximize understandability so that even a novice can find relations easily between computational logic and human daily tasks.

## Lexical Aspects

The lexical aspects of the Humanly are as follows:

### Rules for Identifiers

The rules for identifiers that will be used are as follows:

**Rule 1.** Humanly identifiers will be case sensitive. For example, '*name*' and '*Name*' will be considered as two different identifiers in Humanly.

**Rule 2.** Humanly identifiers can contain only alphanumeric characters ([a-z], [A-Z], [0-9]) and underscore (\_).

**Rule 3.** The first character of an identifier can only contain an alphabet ([a-z], [A-Z]).

**Rule 4.** Identifiers should not start with digits ([0-9]). For example, "*123humanly*" or "*123\_humanly*" is an invalid identifier.

**Rule 5.** There should not be any whitespace in an identifier. For example, "*num ber*" or "*num 1*" is an invalid identifier.

**Rule 6.** Any kind of special characters, such as a *semicolon* (;), *period*(.), *whitespaces* ( ), *slash* (\), or *comma*(,) are not permitted to be used in or as an Identifier. For example, "*human-ly*" is an invalid Humanly identifier rather you can use "*human\_ly*".

**Rule 7.** Reserved Words can't be used as an identifier. For example, "*NUM REPEAT= 20;*" is an invalid statement as '*REPEAT*' is a reserved word.

**Rule 8.** Humanly identifiers cannot contain only digits ([0-9]). For example, "*NUM 20 = 10*" is an invalid statement.

**Rule 9.** There will be no limit on the length of the identifier but it is advisable to use an optimum length of **4 – 15** letters only (for further clarification on Naming Convention refer to [Humanly Best Practices/Naming Conventions](#)).

## Constants

### Integer & Decimal Constants

An integer constant is a sequence of one or more decimal digits ([0-9]). There are no negative integer constants; negative numbers may be obtained by negating an integer constant using the unary - operator. If a variable of type *NUM* is assigned to a constant, it can accept an integer constant. Whereas If a variable of type *DEC* is assigned to a constant, it can accept an integer constant with floating/decimal point.

### String Constant

A string constant is a sequence of zero or more printable characters, spaces, or escape sequences surrounded by double quotes ". Each escape sequence starts with a backslash \ and stands for some sequence of characters. The escape sequences are as follows:

Escape Sequence	Meaning
\n	Newline
\t	Tab
\"	Double quote
\\	Backslash

## Operators and Punctuators

- The binary operators are + - \* / ~= = <- <> < > <= >= AND OR.
- (), {}, [], ",", "." and ";" are used as punctuators.
- A delimiter ";" is used to indicate end of line.
- A dot "." is used to represent a floating-point number. For e.g. (12.5).
- Parentheses groups the expressions in the usual way.
- A leading minus sign negates an integer expression.
- The binary operators +, -, \*, and / require integer operands and return an integer result.
- The binary operators >, <, >= and <= compare their operands, which may be either both integer or both string and produce the integer 1 if the comparison is true and 0 otherwise.
- The binary operators =, ~= and <> can compare any two operands of the same type and return either integer 0 or 1. Integers are the same if they have the same value. Strings are the same if they contain the same characters.

## Comments

Anything that may appear between the comments will be ignored by the compiler. A comment in Humanly will begin with **/#** and ends with **#/**.

### For Example:

```
NUM counter = 0; /# This is a counter to count the loop iterations #/
```

In the above example Humanly compiler will ignore the sentence written between **/#** **#/**.

## Reserved Keywords

The reserved keywords that will be used for Humanly are as follows along with the comparison with a popular programming language C++.

*Note: All the keywords will be case sensitive and will be written in Upper case ([A-Z]) letters only*

Humanly	C++
IF	if
ELSE	else
BUTIF	else if
CHOOSE	switch
LISTEN	cin

SHOW	cout
METHOD	functions
REPEAT	while
COLLECTION	array

### Data Types

Few of the data types that will be allowed in Humanly are as follows along with the comparison with a popular programming language C++.

*Note: All the data types will be case sensitive and will be written in Upper case ([A-Z]) letters only.*

Humanly	C++
NUM	int
WORD	string
DEC	float

### Humanly Best Practices/Naming Conventions

Few of the best practice/naming conventions that can be used are mentioned below:

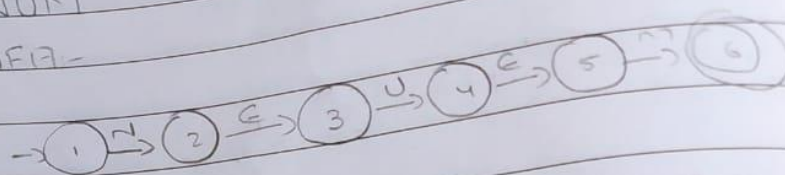
- Use meaningful names that describe the actual data being stored, or the actual activity or meaning implemented by the function or method. (for example: counter can be used if a person is making a counter to count something)
- Names should begin with letters followed by a series of letters, numbers. (For example: num1 and not 1num)
- Names(identifiers/variables) should not start with numbers/digits/underscore. (for e.g.: words like 1num, \_num should not be used)
- Name must be as precise as possible but meaningful. (for e.g.: count for counting)
- Naming convention should also justify rules of programming language such as use either lower casing, Capitalize or only camel casing would do a perfect job. (for e.g.: count, Count or LoopCount)
- Natural language words can also be used to define variable(s) name. (for e.g.: any word that is human understandable such as, count for counting)
- Avoid using special characters in naming conventions. (for e.g.: %, #, \* etc. should not be used)
- Name should avoid using Upper casing as all reserved words and data types for humanly uses upper casing but will not be considered invalid if the identifier is not a keyword. (for e.g.: "NUM NUMBER = 20" is a valid identifier.)

# NFA's, DFA's and Regular Expressions

## Data Types

NUM

NFA -



DFA -

$E\text{-closure}(1) = \{1\} \rightarrow (A)$	$E\text{-closure}(4) = \{4, 5\}$
" $(2) = \{2, 3\}$	" $(5) = \{5\}$
" $(3) = \{3\}$	" $(6) = \{6\}$

move  $(A, N) = E\text{-closure}\{2\} = \{2, 3\} \rightarrow (B)$

"  $(A, U) = \emptyset$

"  $(A, M) = \emptyset$

move  $(B, N) = \emptyset$

"  $(B, U) = E\text{-closure}\{4\} = \{4, 5\} \rightarrow (C)$

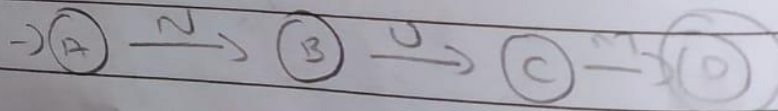
"  $(B, M) = \emptyset$

move  $(C, N) = \emptyset$

"  $(C, U) = \emptyset$

"  $(C, M) = E\text{-closure}\{6\} = \{6\} \rightarrow (D)$

move  $(D) = \emptyset$



Date: \_\_\_\_\_

WORD

NFA



DFA

$\epsilon$ closure $\{1\} = \{1\} \rightarrow (A)$	$\epsilon$ closure $\{5\} = \{5\}$
" $\{2\} = \{2, 3\}$	" $\{6\} = \{6, 7\}$
" $\{3\} = \{3\}$	" $\{7\} = \{7\}$
" $\{4\} = \{4, 5\}$	" $\{8\} = \{8\}$

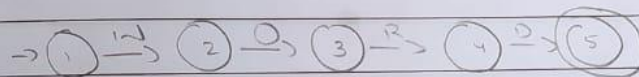
move (A, W) =  $\epsilon$  closure (2) =  $\{2, 3\} \rightarrow (B)$

move (B, D) = " (4) =  $\{4, 5\} \rightarrow (C)$

move (C, R) = " (6) =  $\{6, 7\} \rightarrow (D)$

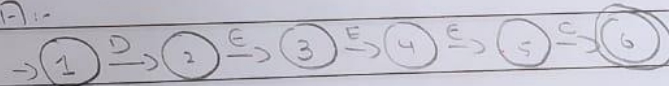
move (D, R) = " (8) =  $\{8\} \rightarrow (E)$

move (E) =  $\phi$



DEC

NFA:-



DFA:-

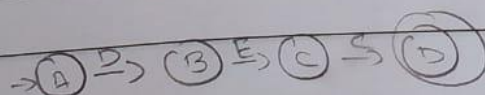
$\epsilon$ closure $\{1\} = \{1\} \rightarrow (A)$	$\epsilon$ closure $\{4\} = \{4, 5\}$
" $\{2\} = \{2, 3\}$	" $\{5\} = \{5\}$
" $\{3\} = \{3\}$	" $\{6\} = \{6\}$

move (A, D) =  $\{2\} = \{2, 3\} \rightarrow (B)$

" (B, E) =  $\{4\} = \{4, 5\} \rightarrow (C)$

" (C, C) =  $\{6\} = \{6\} \rightarrow (D)$

" (D) =  $\phi$





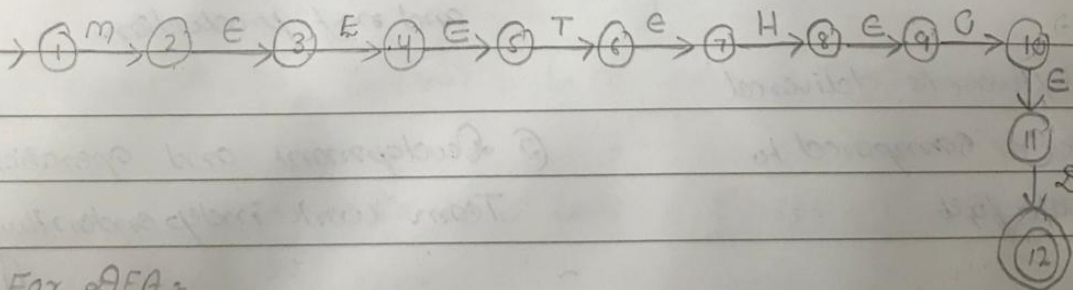
## Reserved Keywords

### NFA and DFA Construction

#### Reserved keywords:-

\* METHOD

For NFA :-



For DFA :-

ECLOSURE (1) = {1} → (A)

// (2) = {2, 3}

(3) = {3}

(4) = {4, 5}

(5) = {5}

ECLOSURE (6) = {6, 7}

// (7) = {7}

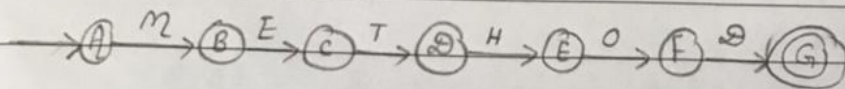
(8) = {8, 9}

(9) = {9}

(10) = {10, 11}

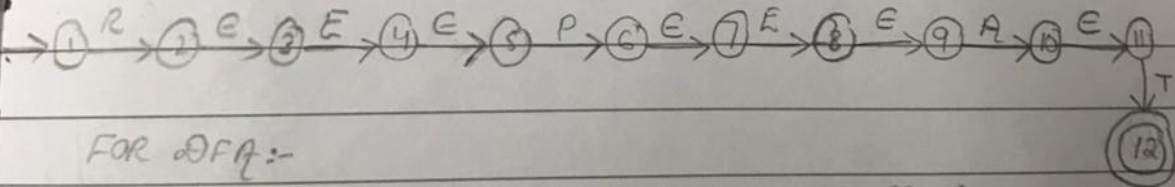


$\text{move } (A, m) = \text{Eclosure } \{2\} = \{2, 3\} \rightarrow (B)$   
 $" (B, E) = \text{Eclosure } \{4\} = \{4, 5\} \rightarrow (C)$   
 $" (C, T) = \text{Eclosure } \{6\} = \{6, 7\} \rightarrow (D)$   
 $" (D, H) = \text{Eclosure } \{8\} = \{8, 9\} \rightarrow (E)$   
 $" (E, O) = \text{Eclosure } \{10\} = \{10, 11\} \rightarrow (F)$   
 $" (F, Q) = \text{Eclosure } \{12\} = \{12\} \rightarrow (G)$   
 $(G) = \emptyset$



\* REPEAT

FOR NFA :-



FOR DFA :-

$\text{Eclosure } (1) = \{1\} \rightarrow (A)$

$\text{Eclosure } (8) = \{8, 9\}$

$" (2) = \{2, 3\}$

$" (9) = \{9\}$

$" (3) = \{3\}$

$" (10) = \{10, 11\}$

$" (4) = \{4, 5\}$

$" (11) = \{11\}$

$" (5) = \{5\}$

$" (12) = \{12\}$

$" (6) = \{6, 7\}$

$" (7) = \{7\}$

Date \_\_\_\_\_

$$\text{Now } (A, R) = \text{Eclosure } \{2\} = \{2, 3\} \rightarrow B.$$

$$" (B, E) = \text{Eclosure } \{4\} = \{4, 5\} \rightarrow C$$

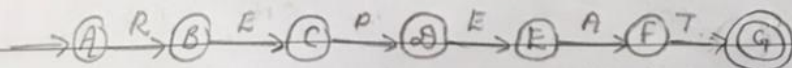
$$" (C, P) = \text{Eclosure } \{6\} = \{6, 7\} \rightarrow D$$

$$" (D, E) = \text{Eclosure } \{8\} = \{8, 9\} \rightarrow E$$

$$" (E, A) = \text{Eclosure } \{10\} = \{10, 11\} \rightarrow F$$

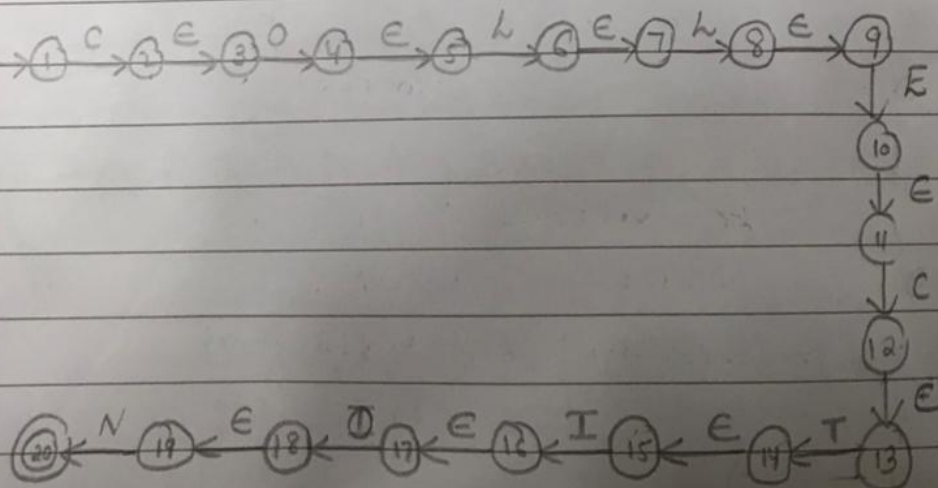
$$" (F, T) = \text{Eclosure } \{12\} = \{12\} \rightarrow G.$$

$$" (G) = \phi$$



\* COLLECTION.

For NFA:-



For DFA:-

Eclosure  $\{1\} = \{1\} \rightarrow (A)$

"  $\{2\} = \{2, 3\}$

"  $\{3\} = \{3\}$

"  $\{4\} = \{4, 5\}$

"  $\{5\} = \{5\}$

"  $\{6\} = \{6, 7\}$

"  $\{7\} = \{7\}$

"  $\{8\} = \{8, 9\}$

"  $\{9\} = \{9\}$

"  $\{10\} = \{10, 11\}$

Eclosure  $\{11\} = \{11\}$

"  $\{12\} = \{12, 13\}$

"  $\{13\} = \{13\}$

"  $\{14\} = \{14, 15\}$

"  $\{15\} = \{15\}$

"  $\{16\} = \{16, 17\}$

"  $\{17\} = \{17\}$

"  $\{18\} = \{18, 19\}$

"  $\{19\} = \{19\}$

"  $\{20\} = \{20\}$

move  $(A, C) = \text{Eclosure } \{2\} = \{2, 3\} \rightarrow B$

"  $(B, O) = \text{Eclosure } \{4\} = \{4, 5\} \rightarrow C$

"  $(C, L) = \text{Eclosure } \{6\} = \{6, 7\} \rightarrow D$

"  $(D, L) = \text{Eclosure } \{8\} = \{8, 9\} \rightarrow E$

"  $(E, E) = \text{Eclosure } \{10\} = \{10, 11\} \rightarrow F$

"  $(F, C) = \text{Eclosure } \{12\} = \{12, 13\} \rightarrow G$

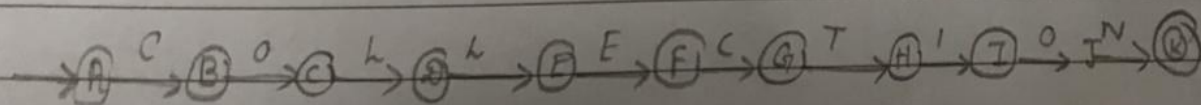
"  $(G, T) = \text{Eclosure } \{14\} = \{14, 15\} \rightarrow H$

"  $(H, I) = \text{Eclosure } \{16\} = \{16, 17\} \rightarrow I$

"  $(I, O) = \text{Eclosure } \{18\} = \{18, 19\} \rightarrow J$

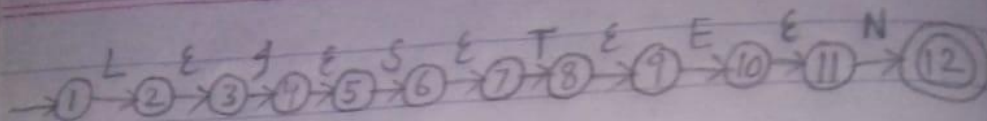
"  $(J, N) = \text{Eclosure } \{20\} = \{20\} \rightarrow K$

"  $(K) = \emptyset$





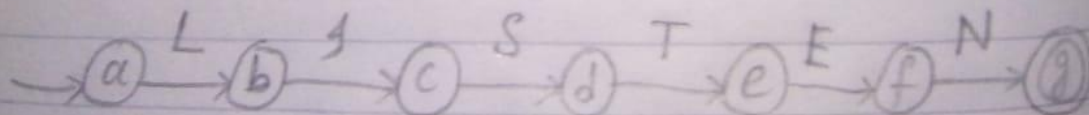
# LASTEN



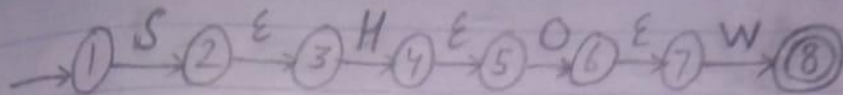
Epsilon Closure:-

$\epsilon$ -closure (1) = {1} = (a)	(7) = {7}
" (2) = {2, 3}	" (8) = {8, 9}
" (3) = {3}	" (9) = {9}
" (4) = {4, 5}	" (10) = {10, 11}
" (5) = {5}	" (11) = {11}
" (6) = {6, 7}	" (12) = {12}

$\epsilon$ -closure (move(a, L)) = $\epsilon$ -closure (2) = {2, 3} = (b)	
" (move(b, J)) = " (4) = {4, 5} = (c)	
" (move(c, S)) = " (6) = {6, 7} = (d)	
" (move(d, T)) = " (8) = {8, 9} = (e)	
" (move(e, E)) = " (10) = {10, 11} = (f)	
" (move(f, N)) = " (12) = {12} = (g)	



SHOW  
~~LISTEN~~



Epsilon Closure:-

$$\epsilon\text{-closure}(1) = \{1\} = \textcircled{a}$$

$$" (2) = \{2, 3\}$$

$$" (3) = \{3\}$$

$$" (4) = \{4, 5\}$$

$$" (5) = \{5\}$$

$$" (6) = \{6, 7\}$$

$$" (7) = \{7\}$$

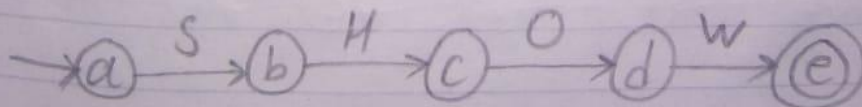
$$" (8) = \{8\}$$

$$\epsilon\text{-closure}(\text{move}(a, S)) = \epsilon\text{-closure}(2) = \{2, 3\} = \textcircled{b}$$

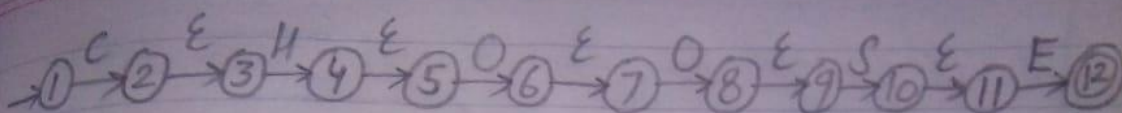
$$" (\text{move}(b, H)) = " (4) = \{4, 5\} = \textcircled{c}$$

$$" (\text{move}(c, O)) = " (6) = \{6, 7\} = \textcircled{d}$$

$$" (\text{move}(d, W)) = " (8) = \{8\} = \textcircled{e}$$



## CHOOSE



Epsilon Closure :-

$$\epsilon\text{-closure}(1) = \{1\} = \textcircled{a}$$

$$\epsilon\text{-closure}(2) = \{2, 3\}$$

$$\epsilon\text{-closure}(3) = \{3\}$$

$$\epsilon\text{-closure}(4) = \{4, 5\}$$

$$\epsilon\text{-closure}(5) = \{5\}$$

$$\epsilon\text{-closure}(6) = \{6, 7\}$$

$$\epsilon\text{-closure}(7) = \{7\}$$

$$\epsilon\text{-closure}(8) = \{8, 9\}$$

$$\epsilon\text{-closure}(9) = \{9\}$$

$$\epsilon\text{-closure}(10) = \{10, 11\}$$

$$\epsilon\text{-closure}(11) = \{11\}$$

$$\epsilon\text{-closure}(12) = \{12\}$$

$$\epsilon\text{-closure}(\text{move}(a, C)) = \epsilon\text{-closure}(2) = \{2, 3\} = \textcircled{b}$$

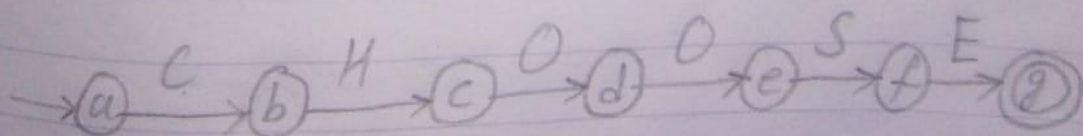
$$\epsilon\text{-closure}(\text{move}(b, H)) = \epsilon\text{-closure}(4) = \{4, 5\} = \textcircled{c}$$

$$\epsilon\text{-closure}(\text{move}(c, O)) = \epsilon\text{-closure}(6) = \{6, 7\} = \textcircled{d}$$

$$\epsilon\text{-closure}(\text{move}(d, O)) = \epsilon\text{-closure}(8) = \{8, 9\} = \textcircled{e}$$

$$\epsilon\text{-closure}(\text{move}(e, S)) = \epsilon\text{-closure}(10) = \{10, 11\} = \textcircled{f}$$

$$\epsilon\text{-closure}(\text{move}(f, E)) = \epsilon\text{-closure}(12) = \{12\} = \textcircled{g}$$





Name: Aheer

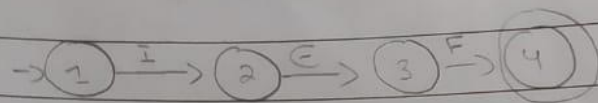
Reg: CS182010

Date:

Reserved Keywords:-

IF

NFA



DFA:-

$E_{closure}(1) = \{1\} \rightarrow (A)$

"  $(2) = \{2, 3\}$

"  $(3) = \{3\}$

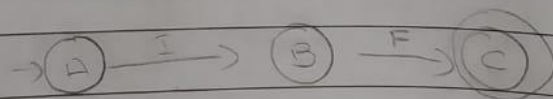
"  $(4) = \{4\}$

$move(A, I) = \{1\}$   $E_{closure}\{2\} = \{2, 3\} \rightarrow (B)$

$move(A, F) = \{1\} = \emptyset$

"  $(B, I) = \{3, 4\} = \emptyset$

"  $(B, F) = \{3, 4\}$   $E_{closure}\{4\} = \{4\} \rightarrow (C)$

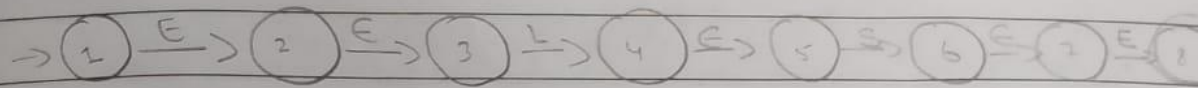




Date: \_\_\_\_\_

ELSE

NFA



DEFA:-

$E_{closure}(1) = \{1\} \rightarrow (A)$	$E_{closure}(5) = \{5\}$
" $(2) = \{2, 3\}$	" $(6) = \{6, 7\}$
" $(3) = \{3\}$	" $(7) = \{7\}$
" $(4) = \{4, 5\}$	" $(8) = \{8\}$

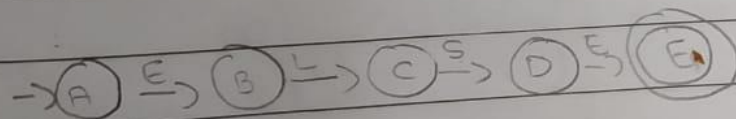
move  $(A, E) = E_{closure}\{2\} = \{2, 3\} \rightarrow (B)$

move  $(B, L) = \{4, 5\} \rightarrow (C)$

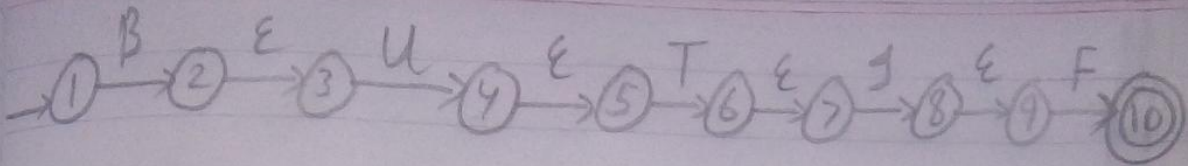
move  $(C, S) = \{6, 7\} \rightarrow (D)$

move  $(D, E) = \{8\} \rightarrow (E)$

move  $(E) = \phi$



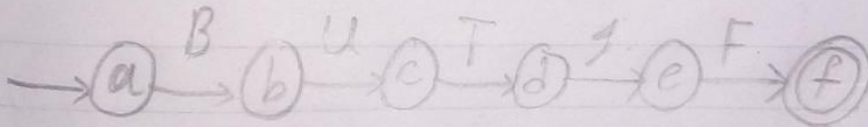
BUT 9F



Epsilon Closure:

$$\begin{array}{lcl}
 \epsilon\text{-closure}(1) = \{1\} = \textcircled{a} & | & \epsilon\text{-closure}(6) = \{6, 7\} \\
 \epsilon\text{-closure}(2) = \{2, 3\} & | & \epsilon\text{-closure}(7) = \{7\} \\
 \epsilon\text{-closure}(3) = \{3\} & | & \epsilon\text{-closure}(8) = \{8, 9\} \\
 \epsilon\text{-closure}(4) = \{4, 5\} & | & \epsilon\text{-closure}(9) = \{9\} \\
 \epsilon\text{-closure}(5) = \{5\} & | & \epsilon\text{-closure}(10) = \{10\}
 \end{array}$$

$$\begin{array}{lcl}
 \epsilon\text{-closure}(\text{move}(a, B)) = \epsilon\text{-closure}(2) = \{2, 3\} = \textcircled{b} \\
 \epsilon\text{-closure}(\text{move}(b, U)) = \epsilon\text{-closure}(4) = \{4, 5\} = \textcircled{c} \\
 \epsilon\text{-closure}(\text{move}(c, T)) = \epsilon\text{-closure}(6) = \{6, 7\} = \textcircled{d} \\
 \epsilon\text{-closure}(\text{move}(d, J)) = \epsilon\text{-closure}(8) = \{8, 9\} = \textcircled{e} \\
 \epsilon\text{-closure}(\text{move}(e, F)) = \epsilon\text{-closure}(10) = \{10\} = \textcircled{f} \\
 \epsilon\text{-closure}(\text{move}(f, \text{ })) = \epsilon\text{-closure}(10) = \{10\} = \textcircled{f}
 \end{array}$$



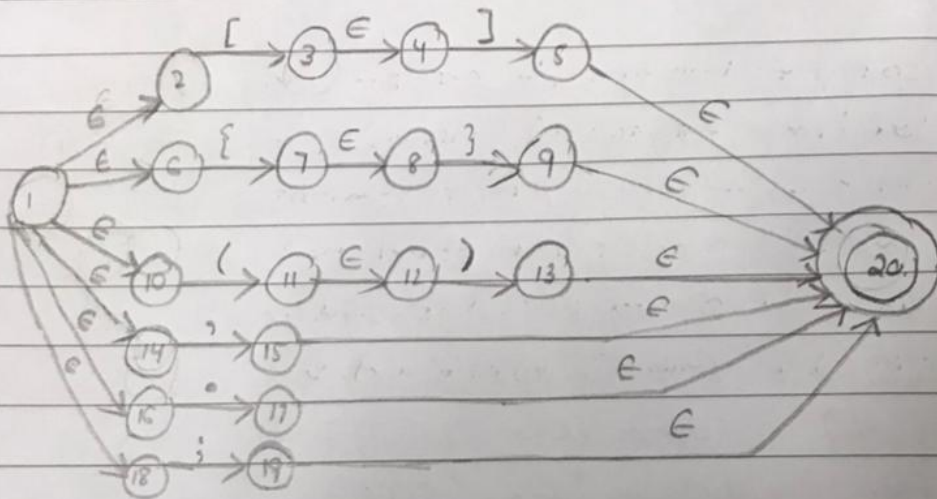
## Punctuators

Date \_\_\_\_\_

Punctuators / Operators:-

\* [ ], { }, ( ), " , " , " . " , " ; "

For NFA:-



For DFA:-

Eclosure {1} = {1}  $\rightarrow A$

{2} = {2}

{3} = {3, 4}

{4} = {4}

{5} = {5, 20}

{6} = {6}

{7} = {7, 8}

Eclosure {8} = {8}

" {9} = {9, 20}

" {10} = {10}

{11} = {11, 12}

{12} = {12}

{13} = {13, 20}

{14} = {14}



Date \_\_\_\_\_

Enclosure {15} = {15, 20}

" {16} = {16}

" {17} = {17, 20}

" {18} = {18}

" {19} = {19, 20}

" {20} = {20}

move (A, [) = Enclosure {3} = {3, 4} → B

(B, ] = Enclosure {5} = {5, 20} → C

(C, { = Enclosure {7} = {7, 8} → D

(D, } = Enclosure {9} = {9, 20} → E

(E, ( = Enclosure {11} = {11, 12} → F

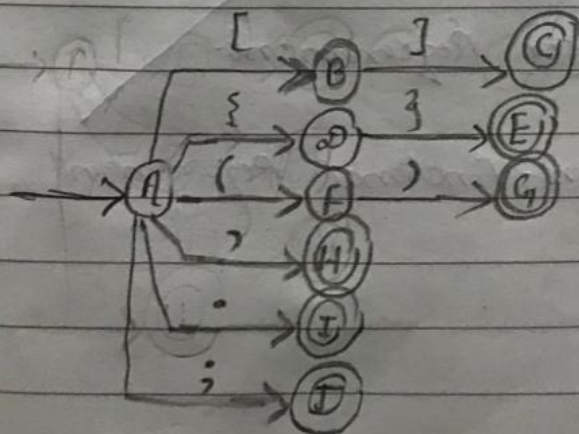
(F, ) = Enclosure {13} = {13, 20} → G

(G, ' = Enclosure {15} = {15, 20} → H

(H, . = Enclosure {17} = {17, 20} → I

(I, ; = Enclosure {19} = {19, 20} → J

(J) =  $\phi$



## Identifiers

Date \_\_\_\_\_

\* Identifiers.

Letter  $[a-zA-Z]$

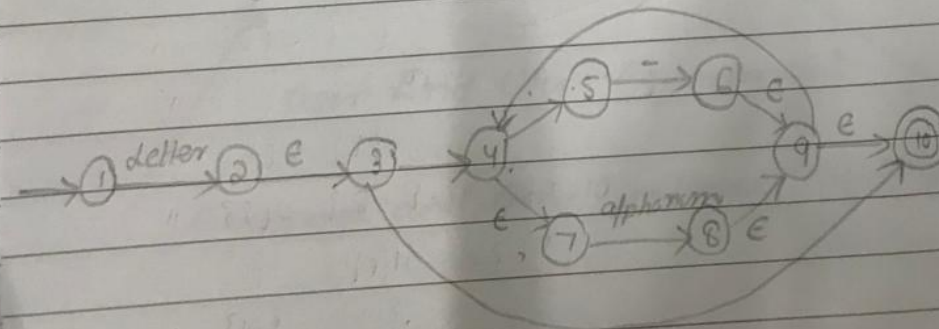
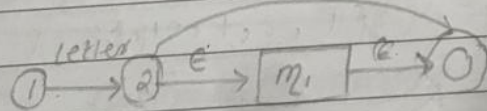
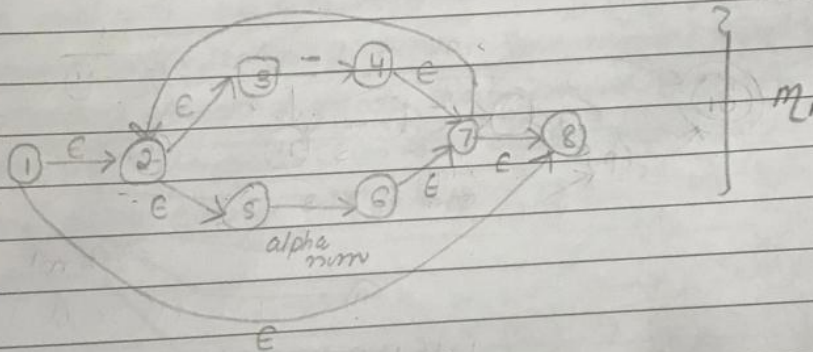
alpha num  $[a-zA-Z0-9]$

Regular Expression:-

$[a-zA-Z][a-zA-Z0-9\_]*$

for NFA:-

$[\_1 \text{ alpha num}]$



Date

Eclosure  $\{1\} = \{1\} \rightarrow \textcircled{A}$

"  $\{2\} = \{2, 3, 4, 5, 7, 10\}$

$\{3\} = \{3, 4, 5, 7, 10\}$

$\{4\} = \{4, 5, 7\}$

$\{5\} = \{5\}$

Eclosure  $\{6\} = \{6, 9, 10, 4, 5, 7\}$

"  $\{7\} = \{7\}$

$\{8\} = \{8, 9, 10, 4, 5, 7\}$

$\{9\} = \{9, 10, 4, 5, 7\}$

$\{10\} = \{10\}$

move  $(A, \text{letter}) = \text{Eclosure } \{2\} = \{2, 3, 4, 5, 7, 10\} \rightarrow \textcircled{B}$

"  $(B, -) = \emptyset$

$(A, \text{alphnum}) = \emptyset$

move  $(B, \text{letter}) = \emptyset$

"  $(B, -) = \text{Eclosure } (6) = \{4, 5, 7, 6, 9, 10\} \rightarrow \textcircled{C}$

$(B, \text{alphnum}) = \text{" } (8) = \{4, 5, 7, 8, 9, 10\} \rightarrow \textcircled{D}$

move  $(C, \text{letter}) = \emptyset$

"  $(C, -) = \text{Eclosure } (6) = \{4, 5, 7, 6, 9, 10\} \rightarrow \textcircled{C}$

"  $(C, \text{alphnum}) = \text{" } (8) = \{4, 5, 7, 8, 9, 10\} \rightarrow \textcircled{D}$

move  $(D, \text{letter}) = \emptyset$

"  $(D, -) = \text{Eclosure } (6) = C$

"  $(D, \text{alphnum}) = \text{" } (8) = D$

