Applications of Finite Automata in Text Search – A Review

Ezhilarasu P[1], Krishnaraj N[2], Suresh Babu V[3]

[1]*Associate Professor/CSE*
Hindusthan College of Engineering and Technology, Coimbatore, India

[2]*Head of the Department/IT*
Sree Sastha Institute of Engineering and Technology, Chennai, India

[3]*Assistant Professor/ECE*
Hindusthan College of Engineering and Technology, Coimbatore, India.

Abstract -- **In this paper, we discuss the text search application of finite automata. First Nondeterministic Finite Automata designed, based on the given keywords. Then Nondeterministic Finite Automata converted into Deterministic Finite Automata. Then given keywords are searched using given paragraph. The transition function used to explain the text search of finite automata.**

Keywords -- **Finite Automata, Text search, Keywords.**

I. INTRODUCTION

Automata theory is the study of abstract machines and automata. It also includes the computational problems that can be solved using them [1]. In the theory of computation, the simpler abstract machine is finite automata. The other important abstract machines are 1. Pushdown Automata 2.Turing Machine. The finite automata proposed to model brain function of the human. The simplest example for finite automata is the switch with two states "on" and "off" [1].

The Finite Automata is the five tuples combination focusing on states and transition through input characters. In figure 1 the ending state is ON, starting state is OFF and collection of states are OFF and ON. It is having only a single input PUSH for making the transition from the state OFF to ON, then ON to OFF. The switch is the simplest practical application of finite automata.
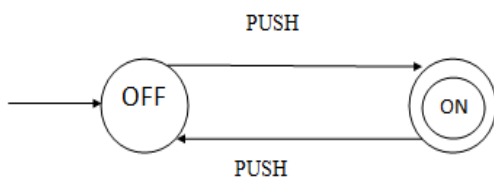
The finite automata concepts also used in numerous fields. Those include lexical analysis, ticket vending machine, pattern recognition, game designing. It also applied to the communication protocol, DNA matching.

II. RELATED WORK

Hopcroft, Motwani and Ullman [2001] listed the applications of finite automata. Finite automata are the useful model for many software and hardware. They used in software for digital circuits, finding text pattern in web pages and verifying systems (Example Communication protocol) [2]. Many research papers [3,4,5,6] and books [7,8,9] published on the application of finite automata.

III. FINITE AUTOMATA APPLICATIONS

Finite Automata concepts used in many applications. In this paper, we discuss text search applications. If, the given problem is to construct the finite automata for accepting the keywords "ezhil", "hills", and "lssbus".

A. CONSTRUCTION OF NONDETERMINISTIC FINITE AUTOMATA

The Nondeterministic Finite Automata can be built quickly for the given problem. It presented as a transition diagram in figure 2 and as a table in Table 1.
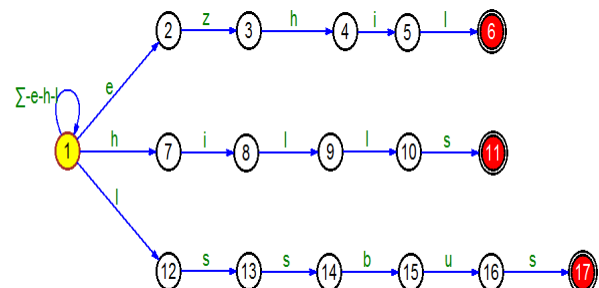

Fig. 1. A Finite Automaton for the switch with on/off states

Fig . 2. Transition Diagram for Nondeterministic Finite Automata for accepting the Keywords "ezhil" "hills" and "lssbus".

Input → States	e	z	h	i	l	s	b	u	Σ-e-z-h-i-l-s-b-u
→1	1,2	1	1,7	1	1,12	1	1	1	1
2		3							
3			4						
4				5					
5					6				
*6									
7				8					
8					9				
9					10				
10						11			
*11									
12						13			
13						14			
14							15		
15								16	
16					17				
*17									

Table . 1. State Transition Table for Nondeterministic Finite Automata for accepting the Keywords "ezhil" "hills" and "lssbus".

Input → States	e	z	h	i	l	s	b	u	Σ-e-z-h-i-l-s-b-u
→1	1,2	1	1,7	1	1,12	1	1	1	1
1,2	1,2	1,3	1,7	1	1,12	1	1	1	1
1,7	1,2	1	1,7	1,8	1,12	1	1	1	1
1,12	1,2	1	1,7	1	1,12	1,13	1	1	1
1,3	1,2	1	1,4,7	1	1,12	1	1	1	1
1,8	1,2	1	1,7	1	1,9,12	1	1	1	1
1,13	1,2	1	1,7	1	1,12	1,14	1	1	1
1,4,7	1,2	1	1,7	1,5,8	1,12	1	1	1	1
1,9,12	1,2	1	1,7	1	1,10,12	1,13	1	1	1
1,14	1,2	1	1,7	1	1,12	1	1,15	1	1
1,5,8	1,2	1	1,7	1	1,6,12	1	1	1	1
1,10,12	1,2	1	1,7	1	1,12	1,11,13	1	1	1
1,15	1,2	1	1,7	1	1,12	1	1	1,16	1
*1,6,12	1,2	1	1,7	1	1,12	1,13	1	1	1
*1,11,13	1,2	1	1,7	1	1,12	1,14	1	1	1
1,16	1,2	1	1,7	1	1,12	1,17	1	1	1
*1,17	1,2	1	1,7	1	1,12	1	1	1	1

Table . 2. State Transition Table for Deterministic Finite Automata for accepting the Keywords "ezhil" "hills" and "lssbus".

B. CONSTRUCTION OF DETERMINISTIC FINITE AUTOMATA

The Deterministic Finite Automata derived from the Nondeterministic Finite Automata. It represented through state transition table as in the table 2.

C. TEXT SEARCH

The input and transition table used for performing the text search. If the given input is "Dr.ezhilarasu umadevi palani went to nilgiri hills by lssbus. ezhil visited many places in that hills. on the way, he saw one flex which has the word ezhillssbus." Then by taking input character by character the output state obtained. The output state 1612 shows that the keyword "ezhil" accepted. The output state 11113 indicates that the keyword "hills" accepted. The output 117 shows that the keyword "lssbus" accepted. The entire processing has given in the table 3. Here the state 1612 obtained as an output state for three times, 11113 obtained as an output state for three times and 117 state for two times. It shows that the words "ezhil" and "hills" occur in the input paragraph for three times. The word "lssbus" for two times, as given in table 3.

S.NO	INPUT STATE	INPUT CHARACTER	OUTPUT STATE	S.NO	INPUT STATE	INPUT CHARACTER	OUTPUT STATE
1.	1	D	1	69.	1	v	1
2.	1	r	1	70.	1	i	1
3.	1	.	1	71.	1	s	1
4.	1	e	12	72.	1	i	1
5.	12	z	13	73.	1	t	1
6.	13	h	147	74.	1	e	12
7.	147	i	158	75.	1	d	1
8.	158	l	**1612**	76.	12		1
9.	1612	a	1	77.	1	m	1
10.	1	r	1	78.	1	a	1
11.	1	a	1	79.	1	n	1
12.	1	s	1	80.	1	y	1
13.	1	u	1	81.	1		1
14.	1		1	82.	1	p	1
15.	1	u	1	83.	1	l	112
16.	1	m	1	84.	112	a	1
17.	1	a	1	85.	1	c	1
18.	1	d	1	86.	1	e	12
19.	1	e	12	87.	12	s	1
20.	12	v	1	88.	1		1
21.	1	i	1	89.	1	i	1
22.	1		1	90.	1	n	1
23.	1	p	1	91.	1		1
24.	1	a	1	92.	1	t	1
25.	1	l	112	93.	1	h	17
26.	112	a	1	94.	17	a	1
27.	1	n	1	95.	1	t	1
28.	1	i	1	96.	1		1
29.	1		1	97.	1	h	17
30.	1	w	1	98.	17	i	18
31.	1	e	12	99.	18	l	1912
32.	12	n	1	100.	1912	l	11012
33.	1	t	1	101.	11012	s	**11113**
34.	1		1	102.	11113	.	1
35.	1	t	1	103.	1		1
36.	1	o	1	104.	1	o	1
37.	1		1	105.	1	n	1
38.	1	n	1	106.	1		1
39.	1	i	1	107.	1	t	1
40.	1	l	112	108.	1	h	17
41.	112	g	1	109.	17	e	12
42.	1	i	1	110.	12		1
43.	1	r	1	111.	1	w	1
44.	1	i	1	112.	1	a	1
45.	1		1	113.	1	y	1
46.	1	h	17	114.	1	,	1
47.	17	i	18	115.	1		1
48.	18	l	1912	116.	1	h	17
49.	1912	l	11012	117.	17	e	12
50.	11012	s	**11113**	118.	12		1
51.	11113		1	119.	1	s	1
52.	1	b	1	120.	1	a	1
53.	1	y	1	121.	1	w	1
54.	1		1	122.	1		1
55.	1	l	112	123.	1	o	1
56.	112	s	113	124.	1	n	1
57.	113	s	114	125.	1	e	12
58.	114	b	115	126.	12		1
59.	115	u	116	127.	1	f	1
60.	116	s	**117**	128.	1	l	112
61.	117	.	1	129.	112	e	12
62.	1		1	130.	1	x	1
63.	1	e	12	131.	1		1
64.	12	z	13	132.	1	w	1
65.	13	h	147	133.	1	h	17
66.	147	i	158	134.	17	i	18
67.	158	l	**1612**	135.	18	c	1
68.	1612		1	136.	1	h	17

S.NO	INPUT STATE	INPUT CHARACTER	OUTPUT STATE
137.	17		1
138.	1	h	17
139.	17	a	1
140.	1	s	1
141.	1		1
142.	1	t	1
143.	1	h	17
144.	17	e	12
145.	12		1
146.	1	w	1
147.	1	o	1
148.	1	r	1
149.	1	d	1
150.	1		1
151.	1	e	12
152.	12	z	13
153.	13	h	147
154.	147	i	158
155.	158	l	**1612**
156.	1612	l	11012
157.	11012	s	**11113**
158.	11113	s	114
159.	114	b	115
160.	115	u	116
161.	116	s	**117**
162.	117	.	1

Table . 3. Text processing for finding keywords

IV. CONCLUSION

This paper is useful for the students, those who are going to study the subject Formal Languages and Automata Theory / Theory of Computation. Students will be able to construct Nondeterministic Finite Automata for the given keywords. Based on the number of keywords and its length the size of the Nondeterministic Finite Automata may vary. The Deterministic Finite Automata possibly constructed from the Nondeterministic Finite Automata.

REFERENCE

[1] http://en.wikipedia.org/wiki/Automata_theory
[2] Hopcraft J E, Motwani R and Ullman J D [2001], "Introduction to Automata Theory, Languages and Computation", Addison-Wesley second edition.
[3] K. Culik II and J. Kari,." Image Compression Using Weighted Finite Automata, in Fractal Image Compression. In Theory a Techniques", Springer-Verlag, pp 243-258, 1994.
[4] Mindek, M., "Finite State Automata and Image Recognition" DATESO 2004, pp 132-143 (2004), ISBN: 80-248-0457-3
[5] G. Navarro, R. Baeza-Yates, "Improving an Algorithm for Approximate String Matching.",Algorithmica, 30(4) 2001
[6] S.V.Ramasubramanian and Kamala Krithivasan,"Finite Automata and Digital Images", IJPRAI, Vol. 14, No. 4, pp. 501-524, 2000.
[7] M. Crochemore, T. Lecroq, "Pattern Matching and Text Compression Algorithms", The Computer Science and Engineering Handbook, A.B. Tucker, Jr, ed., CRC Press, Boca Raton, 2003, Chapter 8.
[8] D. Perrin, "Finite Automata", Handbook of Theoretical Computer Science. Elsevier Science Published 1990.
[9] H. R. Lewis, C. H. Papadimitriou.,"Elements of the Theory of Computation", Prentice Hall 1981.