



**OPPA European Social Fund
Prague & EU: We invest in your future.**

ePAL - Text Searching

Radek Mařík
Marko Genyk-Berezovskyj

ČVUT FEL, K13133

November 21, 2012



- 1 Basic Automata
- 2 Non-deterministic Finite Automaton
- 3 Levenshtein distance
- 4 Dictionary Automata

Outline

- 1 Basic Automata
- 2 Non-deterministic Finite Automaton
- 3 Levenshtein distance
- 4 Dictionary Automata

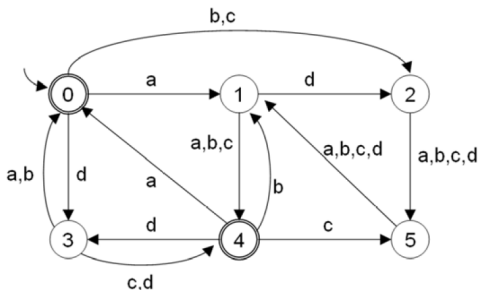
Example 1

Automaton A_1 is given by its transition table. Draw its transition diagram.

	a	b	c	
0	0	1	3	
1	2	2	5	F
2	3	0	2	
3	3	4	1	F
4	1	4	4	
5	5	0	5	

Example 2

Automaton A_2 is given by its transition diagram. Draw its transition table.



Example 3

Make a decision if automaton A_1 accepts the following words

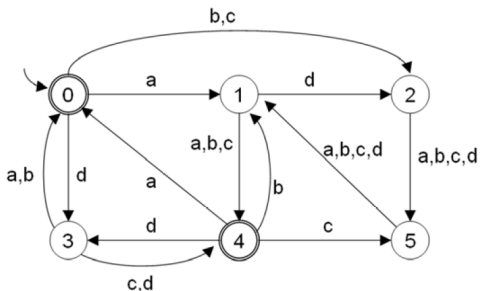
- 1 *addca*
- 2 *bbcca*
- 3 *bbccaba*

	a	b	c	
0	0	1	3	
1	2	2	5	F
2	3	0	2	
3	3	4	1	F
4	1	4	4	
5	5	0	5	

Example 4

Make a decision if automaton A_2 accepts the following words

- 1 *adddca*
- 2 *bbcca*
- 3 *bbccaba*



Example 5

Draw a state diagram of an automaton that accepts just all words over alphabet $\{0, 1\}$ which

- 1 contain subsequence 01,
- 2 do not contain subsequence 01,
- 3 contain a single character 1 and an arbitrary number of characters 0,
- 4 begin and end with symbol 1,
- 5 represent binary representations of numbers 0, 1, 2, 3, 4, 5, 6, 7 in their all 1-, 2- 3- digits sequences.



Outline

- 1 Basic Automata
- 2 Non-deterministic Finite Automaton**
- 3 Levenshtein distance
- 4 Dictionary Automata

Example 6

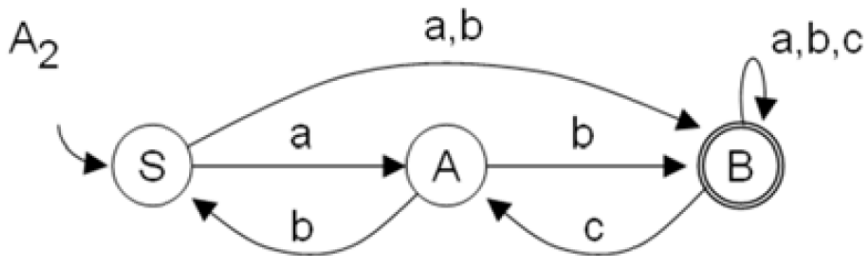
Automaton A_1 is given by its transition table. Determine its equivalent deterministic automaton

A_1

	a	b	c	d	
0	0, 1		2	2	F
1		0, 2			
2	1		1, 2	0, 2	

Example 7

Automaton A_2 is given by its transition table. Determine its equivalent deterministic automaton



Example 8

Create an NFA over alphabet $\{a, b, c\}$ that accepts all words both beginning and ending with chain

- 1 *abc*,
- 2 *acaca*,

Example 9

Create an NFA over alphabet $\{a, b, c\}$ that accepts all words not containing chain

- 1 *abc*,
- 2 *acaca*,



Example 10

Write all words of length at most 5 of a language described by the following regular expression over alphabet $\{0, 1\}$

- 1 $(01|0) \star 0$
- 2 $0(10|0) \star$



Example 11

Write a regular expression describing a maximum set M of words over alphabet $\{a, b, c\}$ such that

- 1 each word in M starts and ends with symbol b ,
- 2 each word in M contains just one occurrence of symbol c anywhere in the word,
- 3 no word in M contains symbol a on an odd position (positions are indexed from 1).



Outline

- 1 Basic Automata
- 2 Non-deterministic Finite Automaton
- 3 Levenshtein distance**
- 4 Dictionary Automata

Example 12

Find all word occurrences in text T having Levenshtein distance at most k from pattern P .

$$T = aacacacbaabbbcbbcacc$$
$$P = abbcba$$
$$k = 2$$


Example 13

Find all word occurrences in text T having Levenshtein distance at most k from pattern P .

$$T = 010011101000010101011100$$
$$P = 11100$$
$$k = 2$$


Outline

- 1 Basic Automata
- 2 Non-deterministic Finite Automaton
- 3 Levenshtein distance
- 4 Dictionary Automata**

Example 14

Create a DFA over alphabet A that accepts just words from set M over this alphabet.

$$A = \{a, b, c\}$$

$$M = \{a, b, ba, bc, aaa, bab, ccc, abbc, abcc\}$$



Example 15

Create a DFA over alphabet A that accepts just words from set M over this alphabet.

$$A = \{0, 1\}$$

$$M = \{10, 11, 101, 111, 1011, 1101, 10001, 10011, 10111, 11101, 11111\}$$



References I



**OPPA European Social Fund
Prague & EU: We invest in your future.**
