

# The Tomita data sets

M. Tomita 1982, “Dynamic construction of finite-state automata from examples using hill-climbing,”  
in Proc. of the Fourth Annual Cognitive Science  
Conf., pp. 105–108

# L1

- POS={1, 11, 111, 1111, 11111, 111111, 1111111, 11111111 }
- NEG={0, 10, 01, 00, 011, 110, 000, 11111110, 10111111 }

# L2

- POS={ 10, 1010, 101010, 10101010,  
101010101010 }
- NEG={ 1, 0, 11, 00, 01, 101, 100, 1001010,  
10110, 11010101 }

L3

- POS={ 1, 0, 01, 11, 00, 100, 110, 111, 000, 100100, 110000011100001, 111101100010011100 }
- NEG={ 10, 101, 010, 1010, 1110, 1011, 10001, 111010, 1001000, 11111000, 0111001101, 11011100110 }

# L4

- POS={ 1, 0, 10, 01, 00, 100100, 11100, 0010, 001111110100, 0100100100 }
- NEG={ 000, 11000, 0001, 0000000000, 11100, 0010, 001111110100, 0000, 11111000011, 101001001, 00000, 1101010000010111 }

# L5

- POS={ 11, 00, 1001, 0101, 1000111101, 1010, 1001100001111010, 111111, 0000 }
- NEG={ 0, 111, 10, 000000000, 1000, 0001, 011, 01, 10, 1110010100, 01011111110 }

# L6

- POS={ 10, 01, 1100, 101010, 111, 000000, 10111, 0111101111, 100100100 }
- NEG={ 1,0,11,00,101,011,11001,1111, 00000000, 010111, 101111011111,1001001001 }

# L7

- POS={ 1, 0, 10, 01, 11111, 000, 00110011, 0101, 0000100001111, 00100, 011111011111, 00 }
- NEG={ 1010, 00110011000, 0101010101, 1011010, 10101, 010100, 101001, 100100110101 }



Table 3. The seven regular languages proposed by Tomita and their respective training sets.

Lan.	Description	Positive instance	Negative instance
1	$1^*$	1, 11, 111, 1111, 11111, 111111, 1111111, 11111111	0, 10, 01, 00, 011, 110, 000, 11111110, 10111111
2	$(1\ 0)^*$	10, 1010, 101010, 10101010, 101010101010	1, 0, 11, 00, 01, 101, 100, 1001010, 10110, 110101010
3	no odd 0-string after an odd 1 string	1, 0, 01, 11, 00, 100, 110, 111, 000, 100100, 110000011100001 , 111101100010011100	10, 101, 010, 1010, 1110, 1011, 10001, 111010, 1001000, 11111000, 0111001101, 11011100110
4	no "000" as a substring	1, 0, 10, 01, 00, 100100, 11100, 0010, 001111110100, 0100100100	000, 11000, 0001, 000000000, 0000, 11111000011, 101001001, 00000, 1101010000010111
5	even number of 0's and even number 1's	11, 00, 1001, 0101, 1000111101, 1010, 1001100001111010, 111111, 0000	0, 111, 10, 000000000, 1000, 01, 10, 1110010100, 010111111110 , 0001, 011
6	(number of 1s - number of 0s) mod 3 = 0	10, 01, 1100, 101010, 111, 000000, 10111, 0111101111 , 100100100	1,0,11,00,101,011,11001, 1111, 00000000, 010111, 101111011111,1001001001
7	$0^*1^*0^*1^*$	1, 0, 10, 01, 1111, 000, 00110011, 0101, 0000100001111, 00100 ,011111011111,00	1010, 00110011000, 0101010101, 1011010, 10101, 010100, 101001, 100100110101