CSE211-Formal Languages and Automata Theory

U4L5_Universal Language Part 1

Dr. P. Saravanan
School of Computing
SASTRA Deemed University

Agenda

- Recap of previous class: Diagonalization
- Converting the Problem to a Language
- Binary-Strings from TM's
- TM Representation by binary
- Language representation by binary
- Universal Language
- Universal Turing Machine

Converting the Problem to a Language

- Let $M = (Q, \Sigma, \Gamma, \delta, q_1, B, \{q_n\})$ be a TM, where

  $Q = \{q_1, q_2, \ldots, q_n\}$, order the states from 1 through n

  $\Sigma = \{x_1, x_2\} = \{0, 1\}$

  $\Gamma = \{x_1, x_2, x_3\} = \{0, 1, B\}$

- Encode each transition:

  $\delta(q_i, x_j) = (q_k, x_l, d_m)$ where q_i and q_k are in ordered Q

x_j and x_l are in Σ, and d_m is in $\{L, R\} = \{d_1, d_2\}$ as:

$$0^i10^j10^k10^l10^m$$

where the number of 0's indicate the corresponding id, and single 1 acts as a barrier

- The TM *M* can then be encoded as:

  $111code_1 11code_2 11code_3 11 \ldots 11code_r 111$

Binary-Strings from TM's

- We shall restrict ourselves to TM's with input alphabet {0, 1}.
- Assign positive integers to the three classes of elements involved in moves:
  1. States: q_1(start state), q_2 (final state), q_3, …
  2. Symbols X_1 (0), X_2 (1), X_3 (blank), X_4, …
  3. Directions D_1 (L) and D_2 (R).

Binary Strings from TM's – (2)

- Less Formally:
  - Every state, tape symbol, and movement symbol is encoded as a sequence of 0's:

q_1,	0	0	0
q_2,	00	1	00
q_3,	000.....	B	000....

L	0
R	00...

Note that 1's are not used to represent the above, since 1 is used as a special separator symbol.

Binary Strings from TM's – (2)

- Suppose $\delta(q_i, X_j) = (q_k, X_l, D_m)$.
- Represent this rule by string $0^i10^j10^k10^l10^m$.
- Key point: since integers i, j, … are all > 0, there cannot be two consecutive 1's in these strings

  - Example:

    $\delta(q_2, 1) = (q_3, 0, R)$ Is encoded as:

    00100100010100

TM Representation by binary

- Represent a TM by concatenating the codes for each of its moves, separated by **11** as punctuation.
  - That is: $Code_1 11 Code_2 11 Code_3 11$ …

	0	1	B
q_1	$(q_1, 0, R)$	$(q_1, 1, R)$	(q_2, B, L)
q_2	$(q_3, 0, R)$	-	-
q_3	-	-	-

What is the L(M)?

Coding for the above table:

1110101010100**11**0100101001001**11**0100010010010**11**00101000101001***111***

Are the followings correct encoding of a TM?

0110000111000 1

111111

Language Representation by binary string

- **Definition:**

$$L_t = \{x \mid x \text{ is in } \{0, 1\}^* \text{ and } x \text{ encodes a TM}\}$$

- Question: Is L_t recursive?
- Answer: Yes. [Check only for format, i.e. the order and number of 0's and 1's, syntax checking]

- Question: Is L_t decidable:
- Answer: Yes (same question).

References

John E. Hopcroft, Rajeev Motwani and Jeffrey D. Ullman, *Introduction to Automata Theory*, Languages, and Computation, Pearson, 3rd Edition, 2011.

Peter Linz, An Introduction to Formal Languages and Automata, Jones and Bartle Learning International, United Kingdom, 6th Edition, 2016.

Next Class: Unit IV

Universal Language Part 2

Thank you.