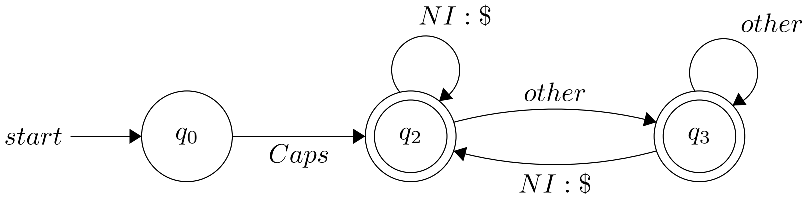
11-611 Natural Language Processing
Homework 2

Jennifer Li

February 2019

1 Transducer-1

Keep the first letter of the name, and drop all occurrences of non-initial a, e, h, i, o, u, w, y

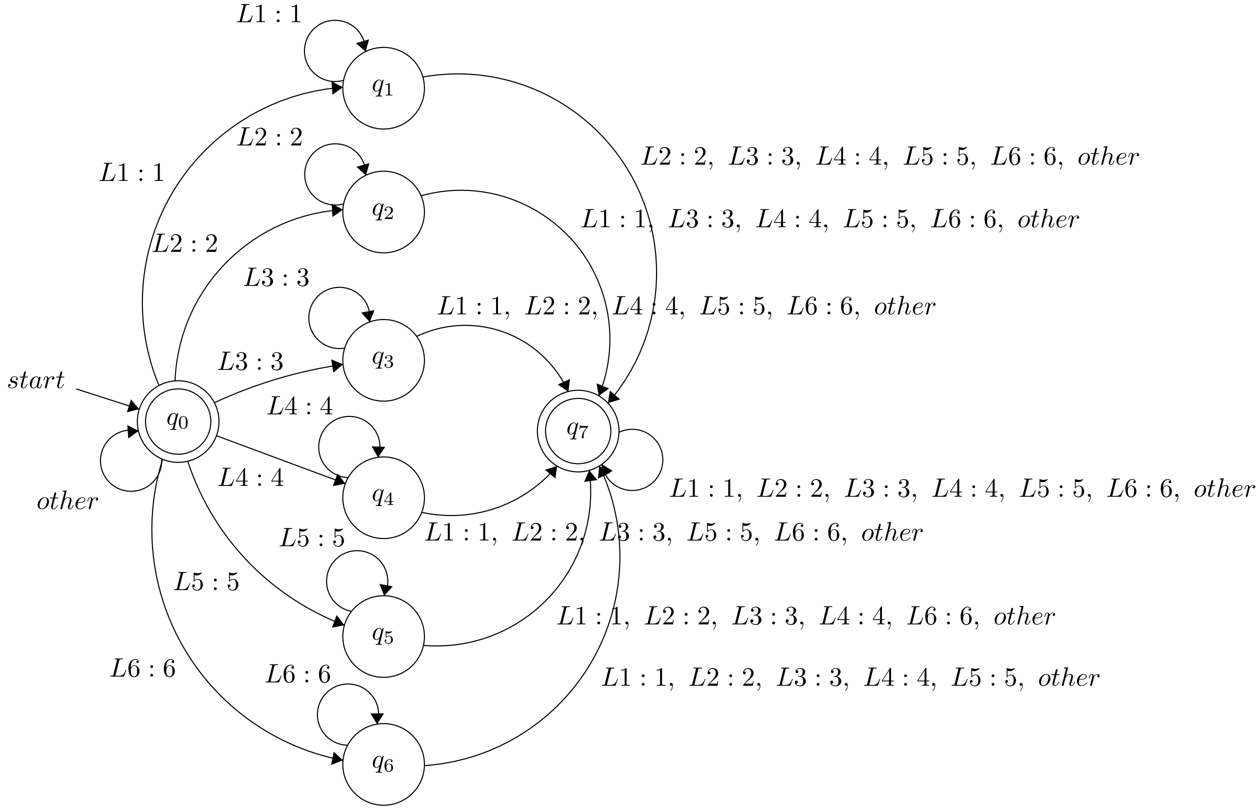


Caps = All capital letters
NI = a, e, h, i, o, u, w, y

2 Transducer-2

Replace the remaining letters with the following numbers:

(a) b, f, p, v replaced with 1
(b) c, g, j, k, q, s, x, z replaced with 2
(c) d, t replaced with 3
(d) l replaced with 4
(e) m, n replaced with 5
(f) r replaced with 6


L1 : 1
q_1
L1 : 1
L2 : 2, L3 : 3, L4 : 4, L5 : 5, L6 : 6, other
L2 : 2
q_2
L2 : 2
L1 : 1, L3 : 3, L4 : 4, L5 : 5, L6 : 6, other
L3 : 3
L1 : 1, L2 : 2, L4 : 4, L5 : 5, L6 : 6, other
q_3
L3 : 3
start
q_0
L4 : 4
q_7
other
L4 : 4
q_4
L1 : 1, L2 : 2, L3 : 3, L4 : 4, L5 : 5, L6 : 6, other
L1 : 1, L2 : 2, L3 : 3, L5 : 5, L6 : 6, other
L5 : 5
L5 : 5
q_5
L1 : 1, L2 : 2, L3 : 3, L4 : 4, L6 : 6, other
L6 : 6
L6 : 6
L1 : 1, L2 : 2, L3 : 3, L4 : 4, L5 : 5, other
q_6

L1 = b, f, p, v

L2 = c, g, j, k, q, s, x, z

L3 = d, t

L4 = l

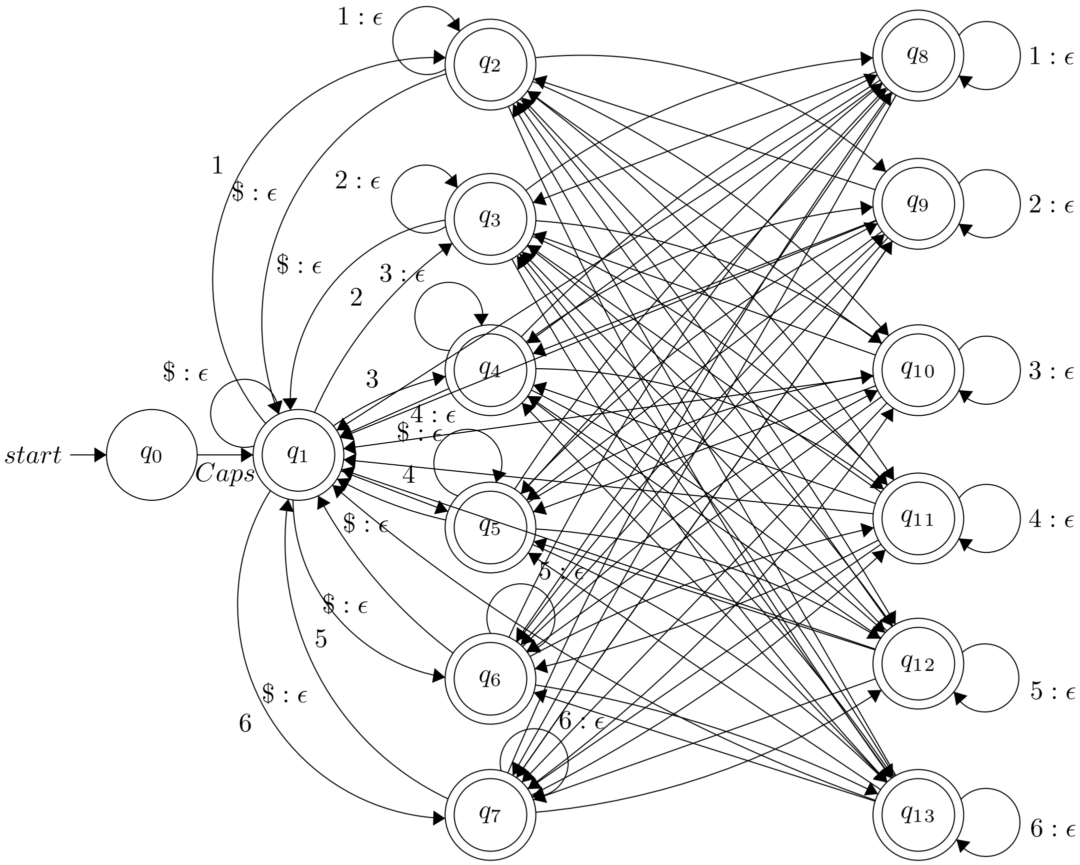
L5 = m, n

L6 = r

⋆ Here I am using ”other” for not changing those place holders($ sign in FST 1) or non-initals (a, e, i, o, u, w, y) that potentially exist

3 Transducer-3

Replace any sequences of identical numbers with a single number, only if they derive from two or more letters that were adjacent in the original name

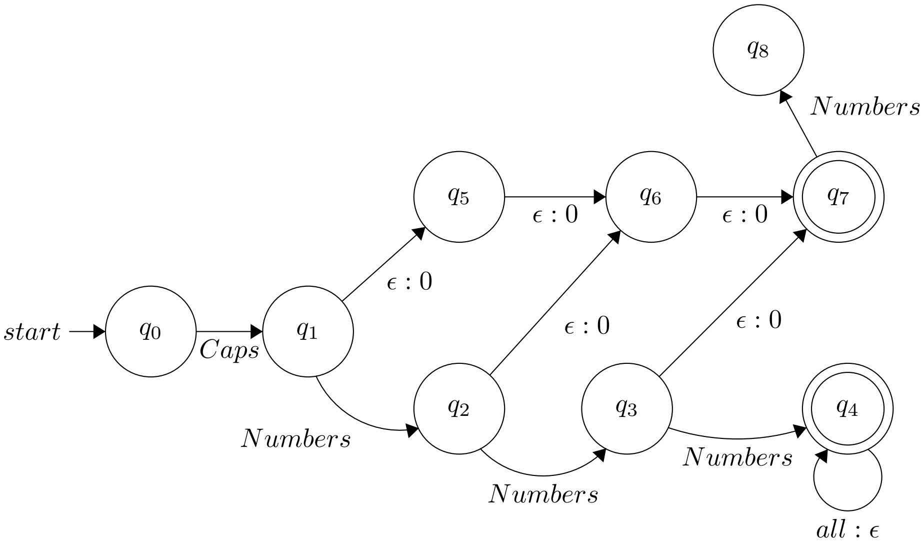


Transition	input:output
q_8, q_9, q_{10}, q_{11}, q_{12}, $q_{13} \to q_1$	$: ϵ
q_3, q_4, q_5, q_6, $q_7 \to q_8$	1
$q_8 \to q_8$	1 : ϵ
q_2, q_4, q_5, q_6, $q_7 \to q_9$	2
$q_9 \to q_9$	2 : ϵ
q_2, q_3, q_5, q_6, $q_7 \to q_{10}$	3
$q_{10} \to q_{10}$	4 : ϵ
q_2, q_3, q_4, q_6, $q_7 \to q_{11}$	4
$q_{11} \to q_{11}$	4 : ϵ
q_2, q_3, q_4, q_5, $q_7 \to q_{12}$	5
$q_{12} \to q_{12}$	5 : ϵ
q_2, q_3, q_4, q_5, $q_6 \to q_{13}$	6
$q_{13} \to q_{13}$	6 : ϵ
$q_8 \to q_3$, q_4, q_5, q_6, q_7	2, 3, 4, 5, 6
$q_9 \to q_2$, q_4, q_5, q_6, q_7	1, 3, 4, 5, 6
$q_{10} \to q_2$, q_3, q_5, q_6, q_7	1, 2, 4, 5, 6
$q_{11} \to q_2$, q_3, q_4, q_6, q_7	1, 2, 3, 5, 6
$q_{12} \to q_2$, q_3, q_4, q_5, q_7	1, 2, 3, 4, 6
$q_{13} \to q_2$, q_3, q_4, q_5, q_6	1, 2, 3, 4, 5

Here, all the place holders ($ sign) are removed and consecutive same numbers are truncated

4 Transducer-4

Convert to the form ”Letter Digit Digit Digit” by dropping the rest of the digits. If there are too few digits, pad with sufficient 0 s.



Numbers = 1, 2, 3, 4, 5, 6
Caps = All capital letters

⋆ Here I am using state q_8 for not letting non-terminated string go through the path q_5 through q_7 since that path is for the strings that do not have enough digits

5 Combination

In order to combine the above four FSTs, assuming the first transducer T_1 takes in the input name S and output O_1, and the second transducer T_2 takes in O_1 and output O_2, thrid transducer T_3 takes O_2 and output O_3 and the fourth transducer T_4 takes O_3 and output O_4. This process maps from S to O_4, which is composition of T_1, T_2, T_3 and T_4 (T$_1 \circ T_2 \circ T_3 \circ T_4$)