

CPE 372/641 Natural Language Processing

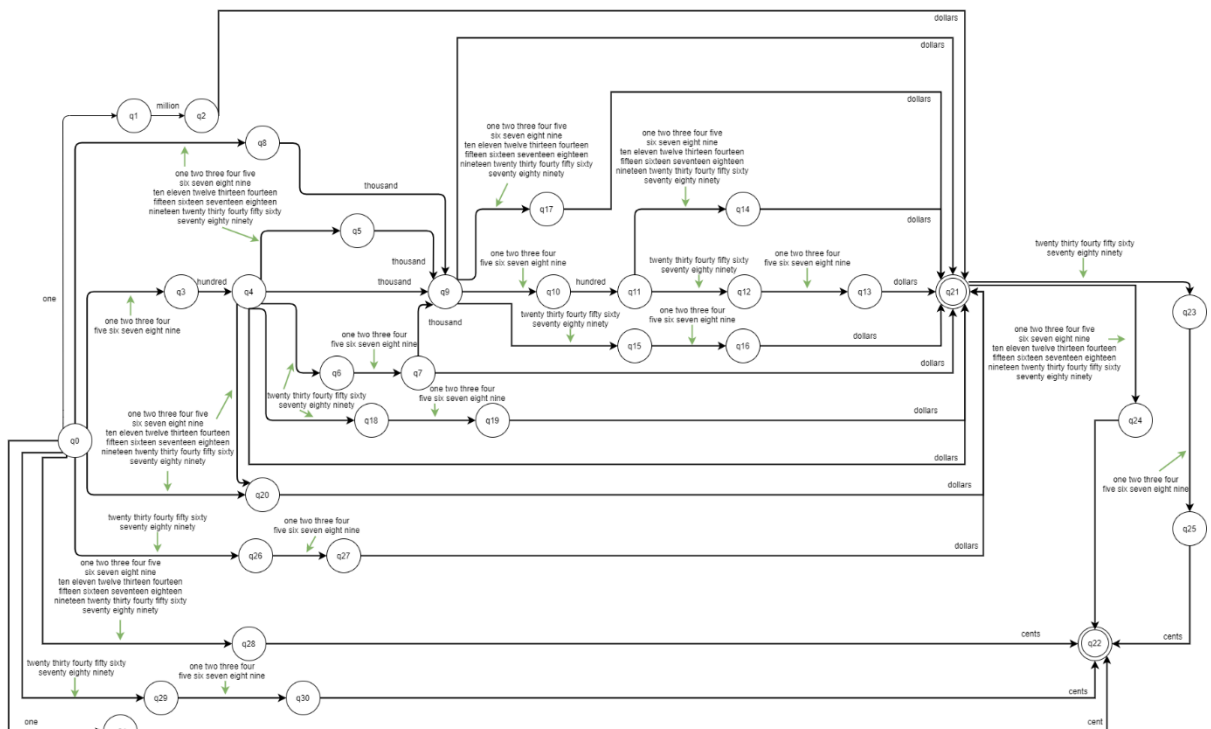
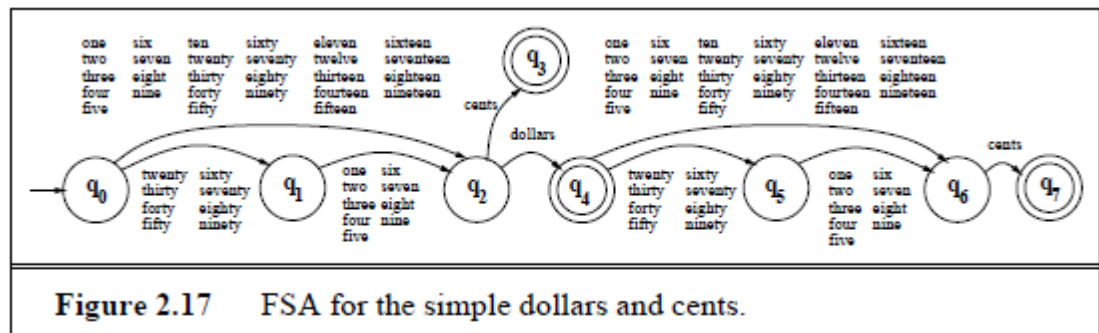
Homework Assignment 1

Regular Expression, Finite State Automata and Minimum Edit Distance

Due on February 8, 2020 via Piazza

(please name your homework pdf file "hw1xxxxxxxxxx(yourid)")

- From the English money expressions that we saw below, complete the FSA to handle the amount up to \$ 1,000,000. Make sure that "dollar" and "cent" have the proper plural endings when appropriate (6 points).



Link for FSA : <https://drive.google.com/file/d/1HyKP88EpC-Z6qJnXtFPu7m6PhjCo1p3/view?usp=sharing>

2. Based on what we learn about minimum edit distance, compute the minimum edit distance matrix for OSLO – SNOW cost: insert=1, delete=1, substitute=1 (4 points).

	0	1	2	3	4	
		S	N	O	W	
0	0	1	2	3	4	$(0,0) \text{ med}('','') = 0$
1	0	1	2	2	3	$(1,0) \text{ med}('s','') = 0+1 = 1$
2	S	2	1	2	3	$(2,0) \text{ med}('s','n') = 1+1 = 2$
3	L	3	2	2	3	$(3,0) \text{ med}('s','no') = 2+1 = 3$
4	O	4	3	3	2	$(4,0) \text{ med}('s','now') = 3+1 = 4$

$(0,1) \text{ med}('','o') = 0+1 = 1$
 $(0,2) \text{ med}('','os') = 1+1 = 2$
 $(0,3) \text{ med}('','osl') = 2+1 = 3$
 $(0,4) \text{ med}('','oslc') = 3+1 = 4$

$(1,1) \text{ med}('s','o') = ?$ $\uparrow (1,0) \text{ med}('s','') = 1+1 = 2$ $\leftarrow (0,1) \text{ med}('','o') = 1+1 = 2$ $\nwarrow (0,0) \text{ med}('','') = 0+1 = 1$	$(4,1) \text{ med}('snow','o') = ?$ $\uparrow (4,0) \text{ med}('snow','') = 4+1 = 5$ $\leftarrow (3,1) \text{ med}('sno','o') = 2+1 = 3$ $\nwarrow (3,0) \text{ med}('sno','') = 3+1 = 4$
$(2,1) \text{ med}('sn','o') = ?$ $\uparrow (2,0) \text{ med}('sn','') = 2+1 = 3$ $\leftarrow (1,1) \text{ med}('s','o') = 1+1 = 2$ $\nwarrow (1,0) \text{ med}('s','') = 1+1 = 2$	$(1,2) \text{ med}('s','os') = ?$ $\uparrow (1,1) \text{ med}('s','o') = 1+1 = 2$ $\leftarrow (0,2) \text{ med}('','os') = 2+1 = 3$ $\nwarrow (0,1) \text{ med}('','o') = 1+0 = 1$

$(3,1) \text{ med}('sno','o') = ?$ $\uparrow (3,0) \text{ med}('sno','') = 3+1 = 4$ $\leftarrow (2,1) \text{ med}('sn','o') = 2+1 = 3$ $\nwarrow (2,0) \text{ med}('sn','') = 2+0 = 2$	$(1,3) \text{ med}('s','oslc') = ?$ $\uparrow (1,2) \text{ med}('s','os') = 1+1 = 2$ $\leftarrow (0,3) \text{ med}('','oslc') = 3+1 = 4$ $\nwarrow (0,2) \text{ med}('','os') = 2+1 = 3$
---	---

$(1,4) \text{ med}('s', 'oslo') = ?$ $\uparrow (1,4) \text{ med}('s', 'osl+o') = 2+1 = \textcircled{3}$ $\leftarrow (0,4) \text{ med}(''+s', 'oslo') = 4+1 = 5$ $\nwarrow (0,3) \text{ med}(''+s', 'osl+o') = 3+1 = 4$	$(4,2) \text{ med}('snow', 'os') = ?$ $\uparrow (4,1) \text{ med}('snow', 'o'+s') = 3+1 = 4$ $\leftarrow (3,2) \text{ med}('sno'+w, 'os') = 3+1 = 4$ $\nwarrow (3,1) \text{ med}('sno'+w, 'o'+s') = 2+1 = 3$
$(2,2) \text{ med}('sn', 'os') = ?$ $\uparrow (2,1) \text{ med}('sn', 'o'+s') = 2+1 = 3$ $\leftarrow (1,2) \text{ med}('s'+n, 'os') = 1+1 = \textcircled{2}$ $\nwarrow (1,1) \text{ med}('s'+n, 'o'+s') = 1+1 = \textcircled{2}$	$(5,3) \text{ med}('snow', 'osl') = ?$ $\uparrow (3,2) \text{ med}('sno', 'os'+l') = 3+1 = 4$ $\leftarrow (2,3) \text{ med}('sn'+o, 'osl') = 2+1 = 3$ $\nwarrow (2,2) \text{ med}('sn'+o, 'os'+l') = 2+1 = 3$

$(2,3) \text{ med}('sn', 'osl') = ?$ $\uparrow (2,2) \text{ med}('sn', 'os'+l') = 2+1 = 3$ $\leftarrow (1,3) \text{ med}('s'+n, 'osl') = 2+1 = 3$ $\nwarrow (1,2) \text{ med}('s'+n, 'os'+l') = 1+1 = \textcircled{2}$	$(3,4) \text{ med}('sno', 'oslo') = ?$ $\uparrow (3,3) \text{ med}('sno', 'osl+o') = 3+1 = 4$ $\leftarrow (2,4) \text{ med}('sn'+o, 'oslo') = 3+1 = 3$ $\nwarrow (2,3) \text{ med}('sn'+o, 'osl+o') = 2+0 = \textcircled{2}$
$(2,4) \text{ med}('sn', 'oslo') = ?$ $\uparrow (2,3) \text{ med}('sn', 'osl+o') = 2+1 = \textcircled{3}$ $\leftarrow (1,4) \text{ med}('s'+n, 'oslo') = 3+1 = 4$ $\nwarrow (1,3) \text{ med}('s'+n, 'osl+o') = 2+1 = \textcircled{3}$	$(4,3) \text{ med}('snow', 'osl') = ?$ $\uparrow (4,2) \text{ med}('snow', 'os'+l') = 3+1 = 4$ $\leftarrow (3,3) \text{ med}('sno'+w, 'osl') = 3+1 = 4$ $\nwarrow (3,2) \text{ med}('sno'+w, 'os'+l') = 3+1 = 4$

$(3,2) \text{ med}('sno', 'os') = ?$ $\uparrow (3,1) \text{ med}('sno', 'o'+s') = 2+1 = 3$ $\leftarrow (2,2) \text{ med}('sn'+o, 'os') = 2+1 = 3$ $\nwarrow (2,1) \text{ med}('sn'+o, 'o'+s') = 2+1 = 3$	$(4,4) \text{ med}('snow', 'oslo') = ?$ $\uparrow (4,3) \text{ med}('snow', 'osl+o') = 4+1 = 5$ $\leftarrow (3,4) \text{ med}('sno'+w, 'oslo') = 2+1 = \textcircled{3}$ $\nwarrow (3,3) \text{ med}('sno'+w, 'osl+o') = 3+1 = 4$
---	---