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))

1. 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).

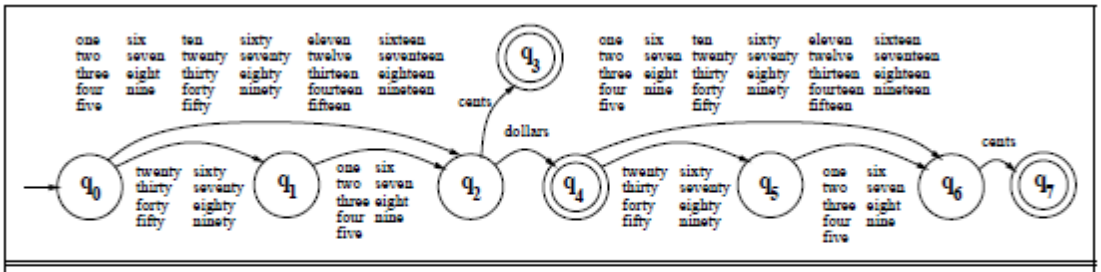
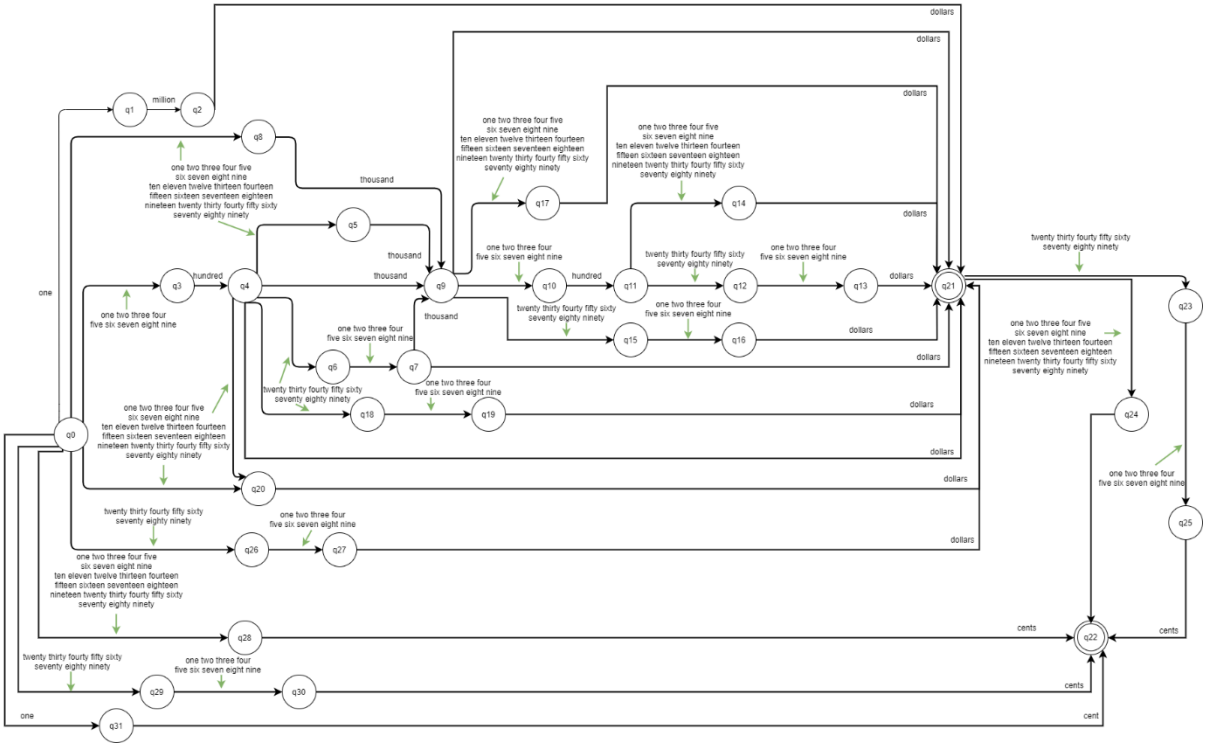
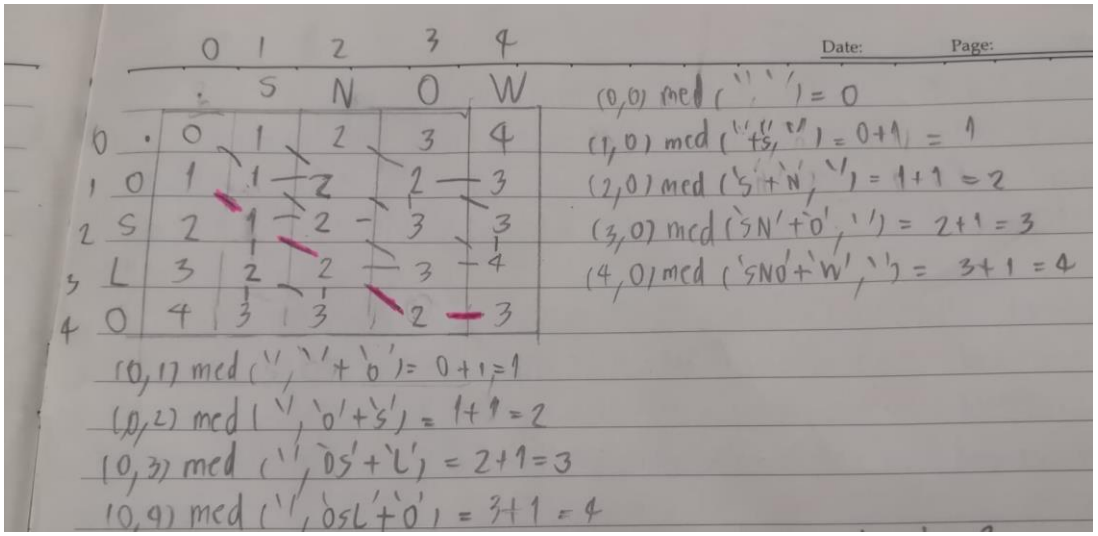


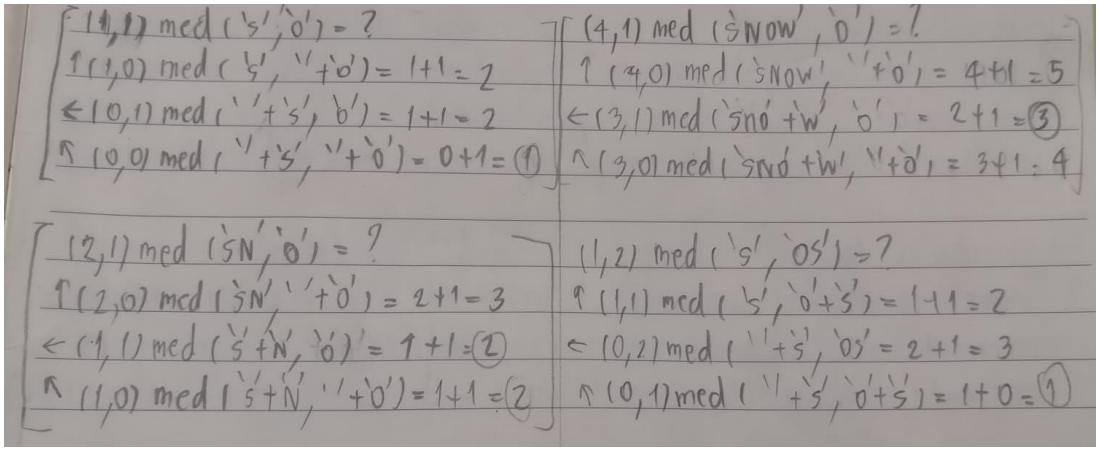
Figure 2.17 FSA for the simple dollars and cents.

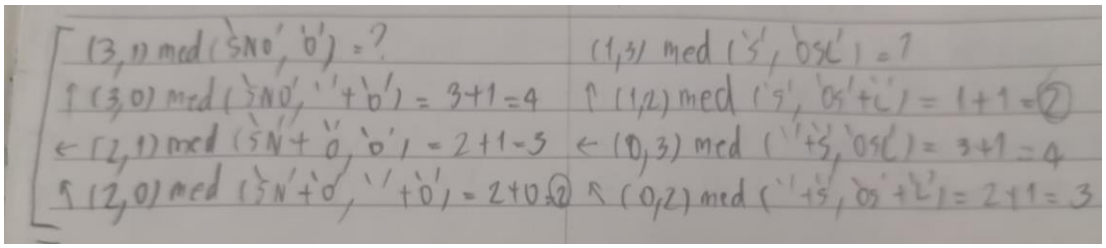


Link for FSA : https://drive.google.com/file/d/1HyKP88EpC-Z6qtJnXtFPu7m6PhjCo1p3/view?usp=sharing

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).







(1,4) med('s', 'osLo') = ?
↑ (1,3) med('s', 'osL'+'o') = 2+1 = 3
← (0,4) med('' + 's', 'osLo') = 4+1 = 5
↖ (0,3) med('' + 's', 'osL' + 'o') = 3+1 = 4

(4,2) med('snow', 'os') = ?
↑ (4,1) med('snow', 'o'+'s') = 3+1 = 4
← (3,2) med('sno' + 'w', 'os') = 3+1 = 4
↖ (3,1) med('sno' + 'w', 'o' + 's') = 2+1 = 3

(2,2) med('sN', 'os') = ?
↑ (2,1) med('sN', 'o' + 's') = 2+1 = 3
← (1,2) med('s' + 'N', 'os') = 1+1 = 2
↖ (1,1) med('s' + 'N', 'o' + 's') = 1+1 = 2

(3,3) med('sNo', 'osL') = ?
↑ (3,2) med('sNo', 'os' + 'L') = 3+1 = 4
← (2,3) med('sN' + 'o', 'osL') = 2+1 = 3
↖ (2,2) med('sN' + 'o', 'os' + 'L') = 2+1 = 3

(2,3) med('sN', 'osL') = ?
↑ (2,2) med('sN', 'os' + 'L') = 2+1 = 3
← (1,3) med('s' + 'N', 'osL') = 2+1 = 3
↖ (1,2) med('s' + 'N', 'os' + 'L') = 1+1 = 2

(3,4) med('sNo', 'osLo') = ?
↑ (3,3) med('sNo', 'osL' + 'o') = 3+1 = 4
← (2,4) med('sN' + 'o', 'osLo') = 3+1 = 3
↖ (2,3) med('sN' + 'o', 'osL' + 'o') = 2+0 = 2

(2,4) med('sN', 'osLo') = ?
↑ (2,3) med('sN', 'osL' + 'o') = 2+1 = 3
← (1,4) med('s' + 'N', 'osLo') = 3+1 = 4
↖ (1,3) med('s' + 'N', 'osL' + 'o') = 2+1 = 3

(4,3) med('snow', 'osL') = ?
↑ (4,2) med('snow', 'os' + 'L') = 3+1 = 4
← (3,3) med('sno' + 'w', 'osL') = 3+1 = 4
↖ (3,2) med('sno' + 'w', 'os' + 'L') = 3+1 = 4

(3,2) med('sNo', 'os') = ?
↑ (3,1) med('sNo', 'o' + 's') = 2+1 = 3
← (2,2) med('sN' + 'o', 'os') = 2+1 = 3
↖ (2,1) med('sN' + 'o', 'o' + 's') = 2+1 = 3

(4,4) med('snow', 'osLo') = ?
↑ (4,3) med('snow', 'osL' + 'o') = 4+1 = 5
← (3,4) med('sno' + 'w', 'osLo') = 2+1 = 3
↖ (3,3) med('sno' + 'w', 'osL' + 'o') = 3+1 = 4