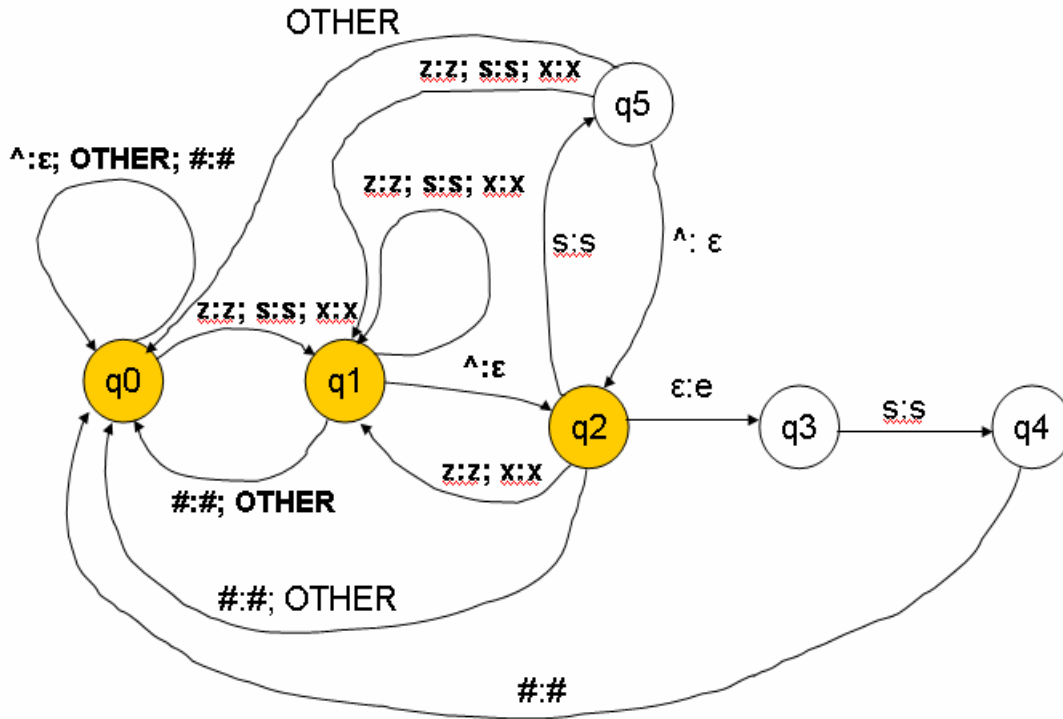


Several Observations on the FST for the E-Insertion Rule

Vladimir Kulyukin
 Department of Computer Science
 Utah State University



FST for the e-insertion rule.

There were a few questions about this FST in class. I checked the original [1] from which I took this example. It has the same FST.

One question was how the FST would handle $FOX^{\wedge}S\#$. On F, we take the OTHER transition, output F, and stay in q_0 ; on O, we take the OTHER transition again, output O, and go back to q_0 ; on X, we transition to q_1 and output X; on \wedge , we transition to q_2 and output ϵ .

When talking about what happens at q_2 , we should bear in mind that this is a nondeterministic FST, because it has transitions on ϵ . We should also remember that when we have reached q_2 , we have processed FOX^{\wedge} and output FOX . We have yet to process $S\#$.

In q_2 , there are 2 choices. We can nondeterministically transition to q_3 and output E or transition on S, output S, and go to q_5 . We cannot go to q_0 on OTHER, because OTHER does not include S or ϵ . If we transition from q_2 to q_5 on S and output S, we will have output $FOX^{\wedge}S$ and still have $\#$ to process in q_5 . Now we are stuck in q_5 , because we

cannot process # since # is not in OTHER. On the other hand, if we transition from q2 to q3 on ϵ and output E, we will have output FOXE, and still have S# to process. From q3 we transition to q4 on S and output S. From q4 we transition to q0 on #, output #, and accept.

If we were to implement this in a program, we would have to either convert this nondeterministic FST to a deterministic one using techniques similar to those used to convert NFAs to DFAs. Or we can use a queue in our processing algorithm to save the multiple choices at each state and then process them systematically, essentially doing a breadth-first search on the graph.

I do not know why they added S:S from q2 to q5. Perhaps, they were trying to handle some exception that they encountered in their text corpus.

References

1. Jurafsky, D. and Martin, J. Speech and Language Processing. Prentice Hall, Upper Saddle River, New Jersey. 2000.