Problem 11

a) T: TEXTESTERTEST(EOF)

Step1: C = (19) , add TE = 28 to the dictionary

Step2: C = (19)(4), add EX = 29 to the dictionary

Step3: C = (19)(4)(23), add XT = 30 to the dictionary

Step4: C = (19)(4)(23)(28), add TES = 31 to the dictionary

Step5: C = (19)(4)(23)(28)(18), add ST = 32 to the dictionary

Step6: C = (19)(4)(23)(28)(18)(28), add TER = 33 to the dictionary

Step7: C = (19)(4)(23)(28)(18)(28)(17), add RT = 34 to the dictionary

Step8: C = (19)(4)(23)(28)(18)(28)(17)(31), add TEST = 35 to the dictionary

Step9: C = (19)(4)(23)(28)(18)(28)(17)(31)(19), add T(EOF) = 36 to the dictionary

Step10: C = (19)(4)(23)(28)(18)(28)(17)(31)(19)(27)

The length of the final encoding info is 60 bits.

b) Compression ratio = $\frac{60 * \log 2}{14 * \log 27} \approx \mathbf{90.1\%}$

The length of C is 60, and alphabet is {1,0}. The length of S is 14, and the alphabet is {Uppercase letter}+{EOF}.

c) When we just added “EMMA” as 342 to the decoding table, and the C tells us that the 343 is the next item. This time, 343 should represents the “EMMAE”. Since the index 343 is just add one step in the encoding process corresponding to the decoding process. So the 343 should be $\mathbf{s_{prev} + s[0]}$