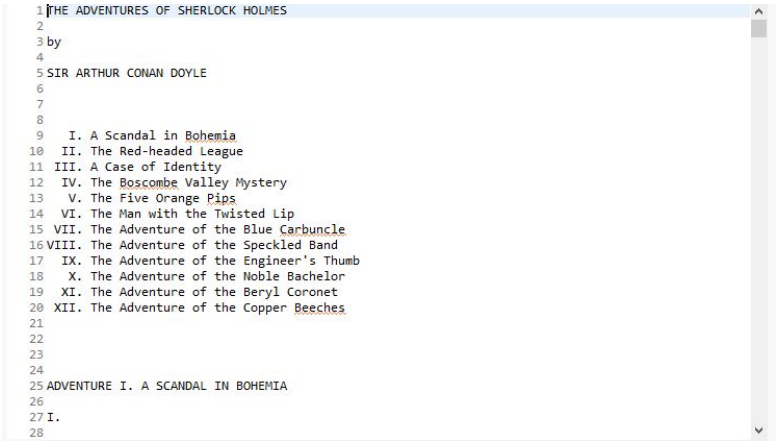
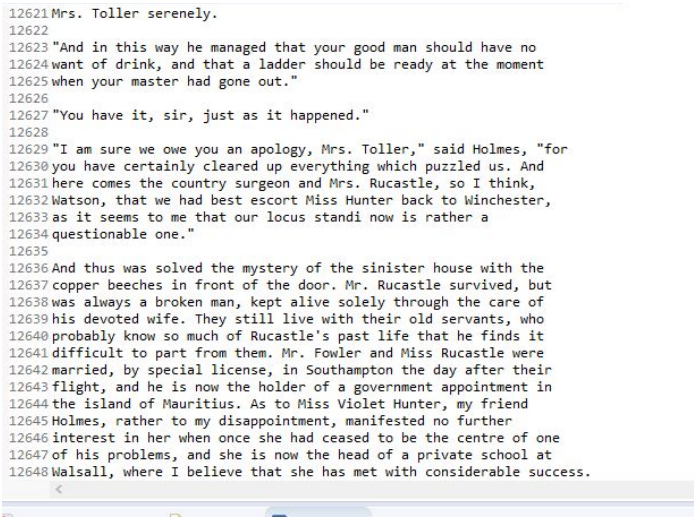
Zachary Werder
720436659

Assignment 1 Write Up

1. Produce text (partial pics for brevity)



```
THE ADVENTURES OF SHERLOCK HOLMES

by

SIR ARTHUR CONAN DOYLE



  I. A Scandal in Bohemia
 II. The Red-headed League
III. A Case of Identity
 IV. The Boscombe Valley Mystery
  V. The Five Orange Pips
 VI. The Man with the Twisted Lip
VII. The Adventure of the Blue Carbuncle
VIII. The Adventure of the Speckled Band
 IX. The Adventure of the Engineer's Thumb
  X. The Adventure of the Noble Bachelor
 XI. The Adventure of the Beryl Coronet
XII. The Adventure of the Copper Beeches



ADVENTURE I. A SCANDAL IN BOHEMIA

I.
```



```
Mrs. Toller serenely.

"And in this way he managed that your good man should have no
want of drink, and that a ladder should be ready at the moment
when your master had gone out."

"You have it, sir, just as it happened."

"I am sure we owe you an apology, Mrs. Toller," said Holmes, "for
you have certainly cleared up everything which puzzled us. And
here comes the country surgeon and Mrs. Rucastle, so I think,
Watson, that we had best escort Miss Hunter back to Winchester,
as it seems to me that our locus standi now is rather a
questionable one."

And thus was solved the mystery of the sinister house with the
copper beeches in front of the door. Mr. Rucastle survived, but
was always a broken man, kept alive solely through the care of
his devoted wife. They still live with their old servants, who
probably know so much of Rucastle's past life that he finds it
difficult to part from them. Mr. Fowler and Miss Rucastle were
married, by special license, in Southampton the day after their
flight, and he is now the holder of a government appointment in
the island of Mauritius. As to Miss Violet Hunter, my friend
Holmes, rather to my disappointment, manifested no further
interest in her when once she had ceased to be the centre of one
of his problems, and she is now the head of a private school at
Walsall, where I believe that she has met with considerable success.
```

Compressed 342 KB
Uncompressed 562 KB

2. Total (Uncompressed) File size: 574992 Symbols
Symbols Probability (rounded to 4 digits)(.00 = symbol not occuring):
(0)through(9)--.00 10 -- .0220 11thr12 --.00 13--.0220 14thr31--.00 32--.1647 33--.0006 34--.0089
35thr37--.00 38--.0000 39--.0026 41--.0000 42thr43--.00 44--.0133 45--.0020 46--.0108 47--.0000
48thr49--.0001 50-- .0000 51thr57--.0000 58--.0001 59--.0003 60thr62--.00 63--.0013 64--.00 65--.0013
66--.0008 67--.0006 68thr69--.0004 70thr71--.0003 72--.0023 73--.0066 74--.0002 75--.0001 76--.0005
77--.0013 78thr79--.0005 80--.0003 81--.0000 82--.0004 83--.0013 84--.0020 85thr86--.0001 87--.0013
88--.0000 89--.0008 90--.0000 91thr96--.00 97--.0598 98--.0102 99--.0178 100--.0319 101--.0919 102--.0153
103--.0134 104--.0484 105--.0456 106--.0006 107--.0060 108--.0293 109--.0192 110--.0494 111--.0577
112--.0116 113--.0007 114--.0423 115--.0460 116--.0658 117--.0227 118--.0076 119--.0183 120--.0009
121--.0156 122--.0003 123thr159--.00 160--.0000 161--.00 162--.0000 163thr167--.00 168thr169--.0000
170thr194--.00 195--.0000 196thr255--.00

3. Theoretical Entropy (base 2)
-(2).022 log .022 + -.1647 log .1647 + -.0006 log .0006 + -.0089 log .0089 + -.0026 log .0026 + -.0133 log .0133 + -(2).002 log .002 + -.0108 log .0108 + -(6).0001 log .0001 + -(5) .0003 log .0003 +-(5) .0013 log .0013 + -(2).0008 log .0008 + -(2).0006 log .0006 + -(3).0004 log .0004 +-.0023 log .0023 + -.0066 log .0066 + -.0002 log .0002 + -(3).0005 log .0005 + -.0598 log .0598 + -.0102 log .0102 + -.0178 log .0178 + -.0319 log .0319 + -.0919 log .0919 + -.0153 log .0153+ -.0134 log .0134 + -.0484 log .0484 + -.0456 log .0456 + -.006 log .006 + -.0293 log .0293 + -.0192 log .0192 + -.0494 log .0494 + -.0577 log .0577 + -.0116 log .0116 + -.0007 log .0007 + -.0423 log .0423 + -.046 log .046 + -.0658 log .0658 + -.0227 log .0227 + -.0076 log .0076 + -.0183 log .0183 + -.0009 log .0009 + -.0156 log .0156

.8431 + .2107 + .0932 + 1.479 + 1.0298 + .8485 = **4.5043 (**Approx**)**

4. Compressed Initial Entropy
Series of Probability times Bit Length found in actual tree: **4.8** (approx)
I had a lot of trouble getting accurate numbers doing this by hand looking through looking at my code, but I understand the process of multiplying the bit lengths times the probabilities collected above. It is clear from comparing trees that the initial compression will be less efficient than the encoders.

5. Recompressed 322 KB
   Reuncompressed 562 KB

6. Compressed Encoder Entropy
Series of Probability times Bit Length found in actual tree: **4.7** (approx)
I had a lot of trouble getting accurate numbers doing this by hand looking through looking at my code, but I understand the process of multiplying the bit lengths times the probabilities collected above. It is clear from comparing trees that the initial compression will be less efficient than the encoders.

7. The encoder achieves better compression entropy for data than the original. (342 to 322 KB)