Megan Rogge
Part 3 of A1

```
symbol 0: p() = 0.0
symbol 1: p() = 0.0
symbol 2: p() = 0.0
symbol 3: p() = 0.0
symbol 4: p() = 0.0
symbol 5: p() = 0.0
symbol 6: p() = 0.0
symbol 7: p() = 0.0
symbol 8: p() = 0.0
symbol 9: p(   ) = 0.0
symbol 10: p(
) = 0.021995088627320032
symbol 11: p(
) = 0.0
symbol 12: p(
```

) = 0.0
symbol 13: p(
) = 0.021995088627320032
symbol 14: p() = 0.0
symbol 15: p() = 0.0
symbol 16: p() = 0.0
symbol 17: p() = 0.0
symbol 18: p() = 0.0
symbol 19: p() = 0.0
symbol 20: p() = 0.0
symbol 21: p() = 0.0
symbol 22: p() = 0.0
symbol 23: p() = 0.0
symbol 24: p() = 0.0
symbol 25: p() = 0.0
symbol 26: p() = 0.0
symbol 27: p() = 0.0
symbol 28: p() = 0.0
symbol 29: p() = 0.0
symbol 30: p() = 0.0
symbol 31: p() = 0.0
symbol 32: p( ) = 0.16466316053092914
symbol 33: p(!) = 6.000083479422323E-4
symbol 34: p(") = 0.0088575145393327 21
symbol 35: p(#) = 0.0
symbol 36: p($) = 0.0
symbol 37: p(%) = 0.0
symbol 38: p(&) = 8.695773158583076E-6
symbol 39: p(') = 0.0025982970197846 23
symbol 40: p(() = 8.695773158583076E-6

symbol 41: p()) = 8.695773158583076E-6
symbol 42: p(*) = 0.0
symbol 43: p(+) = 0.0
symbol 44: p(,) = 0.01328540223168322 3
symbol 45: p(-) = 0.0019878537440520 91
symbol 46: p(.) = 0.01079319364443331 4
symbol 47: p(/) = 1.7391546317166151E-6
symbol 48: p(0) = 1.4261067980076244E-4
symbol 49: p(1) = 1.0956674179814675E-4
symbol 50: p(2) = 6.08704121100815 3E-5
symbol 51: p(3) = 2.6087319475749227E-5
symbol 52: p(4) = 3.8261401897765534E-5
symbol 53: p(5) = 2.7826474107465842E-5
symbol 54: p(6) = 2.4348164844032614E-5
symbol 55: p(7) = 3.130478337089907E-5
symbol 56: p(8) = 6.434872137351476E-5
symbol 57: p(9) = 2.2609010212316E-5
symbol 58: p(:) = 1.043492779029969E-4
symbol 59: p(;) = 3.513092356067563E-4
symbol 60: p(<) = 0.0
symbol 61: p(=) = 0.0
symbol 62: p(>) = 0.0
symbol 63: p(?) = 0.0012817569635751455
symbol 64: p(@) = 0.0
symbol 65: p(A) = 0.0013234966747363442
symbol 66: p(B) = 8.330550685922587E-4
symbol 67: p(C) = 5.687035645713332E-4
symbol 68: p(D) = 3.7217909118735567E-4
symbol 69: p(E) = 4.0870133845340455E-4
symbol 70: p(F) = 3.304393800261569E-4

```
symbol 71: p(G) = 2.678298132843587E-4
symbol 72: p(H) = 0.002163508361855469 4
symbol 73: p(I) = 0.006563569580098506
symbol 74: p(J) = 1.982636280156941 2E-4
symbol 75: p(K) = 1.356540612738959 8E-4
symbol 76: p(L) = 5.06093997829535E-4
symbol 77: p(M) = 0.001273061190416562 3
symbol 78: p(N) = 5.28703008041851E-4
symbol 79: p(O) = 5.426162450955839E-4
symbol 80: p(P) = 3.182652976041406E-4
symbol 81: p(Q) = 3.47830926343323E-5
symbol 82: p(R) = 3.54787544870189 5E-4
symbol 83: p(S) = 0.001325235829368060 7
symbol 84: p(T) = 0.001980897125525225
symbol 85: p(U) = 8.00011130589643E-5
symbol 86: p(V) = 1.478281436959123E-4
symbol 87: p(W) = 0.001318279210841194 2
symbol 88: p(X) = 1.391323705373292 1E-5
symbol 89: p(Y) = 7.913153574310599E-4
symbol 90: p(Z) = 3.478309263433230 3E-6
symbol 91: p([) = 0.0
symbol 92: p(\) = 0.0
symbol 93: p(]) = 0.0
symbol 94: p(^) = 0.0
symbol 95: p(_) = 0.0
symbol 96: p(`) = 0.0
symbol 97: p(a) = 0.059779962155995216
symbol 98: p(b) = 0.010229707543757 13
symbol 99: p(c) = 0.017685463449926 26
symbol 100: p(d) = 0.031910009182736454
```

symbol 101: p(e) = 0.0919299746779085 6
symbol 102: p(f) = 0.0152784734396304 65
symbol 103: p(g) = 0.0134488827670645 86
symbol 104: p(h) = 0.048351977070985 33
symbol 105: p(i) = 0.045851072710576 84
symbol 106: p(j) = 5.878342655202159E-4
symbol 107: p(k) = 0.006026170798898072
symbol 108: p(l) = 0.029309973008320115
symbol 109: p(m) = 0.019224615298995464
symbol 110: p(n) = 0.049351990984222 39
symbol 111: p(o) = 0.057738194618359905
symbol 112: p(p) = 0.011568856610178924
symbol 113: p(q) = 7.060967804769458E-4
symbol 114: p(r) = 0.042266675014608 9
symbol 115: p(s) = 0.045962378607006 71
symbol 116: p(t) = 0.065809611264156 71
symbol 117: p(u) = 0.022729011881904442
symbol 118: p(v) = 0.007594888276706458
symbol 119: p(w) = 0.018273297715446477
symbol 120: p(x) = 9.408826557586889E-4
symbol 121: p(y) = 0.015610651974288337
symbol 122: p(z) = 2.556557308623424E-4
symbol 123: p({) = 0.0
symbol 124: p(|) = 0.0
symbol 125: p(}) = 0.0
symbol 126: p(~) = 0.0
symbol 127: p() = 0.0
symbol 128: p() = 0.0
symbol 129: p() = 0.0
symbol 130: p() = 0.0

```
symbol 131: p() = 2.6087319475749227E-5
symbol 132: p() = 0.0
symbol 133: p() = 0.0
symbol 134: p() = 0.0
symbol 135: p() = 0.0
symbol 136: p() = 0.0
symbol 137: p() = 0.0
symbol 138: p() = 0.0
symbol 139: p() = 0.0
symbol 140: p() = 0.0
symbol 141: p() = 0.0
symbol 142: p() = 0.0
symbol 143: p() = 0.0
symbol 144: p() = 0.0
symbol 145: p() = 0.0
symbol 146: p() = 0.0
symbol 147: p() = 0.0
symbol 148: p() = 0.0
symbol 149: p() = 0.0
symbol 150: p() = 0.0
symbol 151: p() = 0.0
symbol 152: p() = 0.0
symbol 153: p() = 0.0
symbol 154: p() = 0.0
symbol 155: p() = 0.0
symbol 156: p() = 0.0
symbol 157: p() = 0.0
symbol 158: p() = 0.0
symbol 159: p() = 0.0
symbol 160: p( ) = 1.7391546317166151E-6
```

symbol 161: p(¡) = 0.0
symbol 162: p(¢) = 1.7391546317166151E-6
symbol 163: p(£) = 0.0
symbol 164: p(¤) = 0.0
symbol 165: p(¥) = 0.0
symbol 166: p(¦) = 0.0
symbol 167: p(§) = 0.0
symbol 168: p(¨) = 1.7391546317166151E-6
symbol 169: p(©) = 2.0869855580599383E-5
symbol 170: p(ª) = 0.0
symbol 171: p(«) = 0.0
symbol 172: p(¬) = 0.0
symbol 173: p() = 0.0
symbol 174: p(®) = 0.0
symbol 175: p(¯) = 0.0
symbol 176: p(°) = 0.0
symbol 177: p(±) = 0.0
symbol 178: p(²) = 0.0
symbol 179: p(³) = 0.0
symbol 180: p(´) = 0.0
symbol 181: p(µ) = 0.0
symbol 182: p(¶) = 0.0
symbol 183: p(·) = 0.0
symbol 184: p(¸) = 0.0
symbol 185: p(¹) = 0.0
symbol 186: p(º) = 0.0
symbol 187: p(») = 0.0
symbol 188: p(¼) = 0.0
symbol 189: p(½) = 0.0
symbol 190: p(¾) = 0.0

symbol 191: p(¿) = 0.0
symbol 192: p(À) = 0.0
symbol 193: p(Á) = 0.0
symbol 194: p(Â) = 2.6087319475749227E-5
symbol 195: p(Ã) = 2.6087319475749227E-5
symbol 196: p(Ä) = 0.0
symbol 197: p(Å) = 0.0
symbol 198: p(Æ) = 0.0
symbol 199: p(Ç) = 0.0
symbol 200: p(È) = 0.0
symbol 201: p(É) = 0.0
symbol 202: p(Ê) = 0.0
symbol 203: p(Ë) = 0.0
symbol 204: p(Ì) = 0.0
symbol 205: p(Í) = 0.0
symbol 206: p(Î) = 0.0
symbol 207: p(Ï) = 0.0
symbol 208: p(Ð) = 0.0
symbol 209: p(Ñ) = 0.0
symbol 210: p(Ò) = 0.0
symbol 211: p(Ó) = 0.0
symbol 212: p(Ô) = 0.0
symbol 213: p(Õ) = 0.0
symbol 214: p(Ö) = 0.0
symbol 215: p(×) = 0.0
symbol 216: p(Ø) = 0.0
symbol 217: p(Ù) = 0.0
symbol 218: p(Ú) = 0.0
symbol 219: p(Û) = 0.0
symbol 220: p(Ü) = 0.0

symbol 221: p(Ý) = 0.0
symbol 222: p(Þ) = 0.0
symbol 223: p(ß) = 0.0
symbol 224: p(à) = 0.0
symbol 225: p(á) = 0.0
symbol 226: p(â) = 0.0
symbol 227: p(ã) = 0.0
symbol 228: p(ä) = 0.0
symbol 229: p(å) = 0.0
symbol 230: p(æ) = 0.0
symbol 231: p(ç) = 0.0
symbol 232: p(è) = 0.0
symbol 233: p(é) = 0.0
symbol 234: p(ê) = 0.0
symbol 235: p(ë) = 0.0
symbol 236: p(ì) = 0.0
symbol 237: p(í) = 0.0
symbol 238: p(î) = 0.0
symbol 239: p(ï) = 0.0
symbol 240: p(ð) = 0.0
symbol 241: p(ñ) = 0.0
symbol 242: p(ò) = 0.0
symbol 243: p(ó) = 0.0
symbol 244: p(ô) = 0.0
symbol 245: p(õ) = 0.0
symbol 246: p(ö) = 0.0
symbol 247: p(÷) = 0.0
symbol 248: p(ø) = 0.0
symbol 249: p(ù) = 0.0
symbol 250: p(ú) = 0.0

```
symbol 251: p(û) = 0.0
symbol 252: p(ü) = 0.0
symbol 253: p(ý) = 0.0
symbol 254: p(þ) = 0.0
symbol 255: p(ÿ) = 0.0
```

As calculated by my HuffEncode class:

```
double[] probs = new double[256];
for(int i = 0; i < 256; i++) {
    probs[i] = (symbol_counts[i]*1.000000000)/num_symbols;
    System.out.println("symbol "+i+": p("+(char) i+") = " +probs[i]);
}
```

```
theoretical entropy is 4.532337589916681 bits per symbol
```

```
double entropy = 0;
for(int i = 0; i < 256; i++) {
    if(probs[i]!=0) {
    entropy += (probs[i]*(Math.log(probs[i])/Math.log(2)));
}}
System.out.println("theoretical entropy is "+-entropy+" bits per symbol");
```

```
actual entropy is 4.573566240921613 bits per symbol
```

```
double actualEntropy = 0;
for (int i = 0; i < 256; i++) {
    actualEntropy+=encoder.getCode(i).length()*probs[i];

System.out.println("actual entropy is "+actualEntropy+" bits per symbol");
```

The compressed entropy achieved by my encoder is

Size of the compressed file/ number of characters in uncompressed file

Size of the compressed file = 329,000 bytes/ 574,992 symbols
= 0.57218187383 bytes/symbol = 4.57745499068 bits/symbol

It achieves better compression than the original compressed file because the original was 349,484 bytes and this was 329,000 (for the same number of symbols). This makes sense since I implemented a minimum variance encoder and the one used to make the original compressed file wansn’t guaranteed to be so.