Solution to Lab 1

- **We start with computing the self entropy of a letter or a few, *X*, as:**

$$H(X) = -\sum_{x} p(x) \log_2 p(x)$$

- **Then we computed the joint entropy of two consecutive events as:**

$$H(X,Y) = H(X) + H(Y \mid X) = H(Y) + H(X \mid Y) = -\sum_{x,y} p(x,y) \log_2 p(x,y)$$

- **We are interested in the number of bits required to predict a letter, *Y*, given a preceding letter or a letter group, *X*, which is equivalent to calculating the conditional entropy:**

$$H(Y \mid X) = H(X,Y) - H(X)$$

- **Here are the results based on the WSJ data:**

Given	zero	one	two	three
1K sentences	4.13	3.42	2.73	2.00
10K sentences	4.13	3.43	2.77	2.12