- What are the 10 (20, 50, etc) most frequent word lengths?

- Find a word that has three consecutive double letters. To illustrate, the word “committee”, “c-o-m-m-i-t-t-e-e”, almost qualifies, except for the ‘i’. (In other words, if “commttee” was a word, it would be an answer to this question).

- Counting bi-grams.

A bi-gram is a pair of consecutive letters in a word. For example, the word “reverend” contains the bi-grams “re”, “ev”, “ve”, “er”, “re” (again), “en” and “nd”. Bi-gram frequency (or, more generally, n-gram frequency) is used for various kinds of statistical analysis of text, for example in automatic document classification. Questions: –

- What are the 10 (20, 50, etc) most common bi-grams across the entire word list?
- Using the letters of the English alphabet, there are 26 ** 2 possible bi-grams. How many (and which!) of these do not appear in any word?
- How many words contain repeated bi-grams?
- What is the highest number of repetions of any bi-gram in a word, and which words have that number of repetitions?

To find the most frequent words or bi-grams, it is convenient to sort the contents of the dictionary by value (the count). You can build up a list of the key-value pairs by iterating over the items in the dictionary. What happens when you sort this list? A key-value pair is a sequence (of length two). How does python compare sequences? (Hint: You can reorder the two elements of the pair when you construct the list.)