

Module 19 Activities: Dictionaries

1. "Insecure Securer Password Generation": Write a function that takes a string that replaces each occurrence of a vowel with another one. You should use a dictionary to implement the substitution. `subdic = {"a": "o", "e": "i", ...}`. Remember that you can use `if letter in subdic` in order to ascertain whether there is a substitution for a given letter.
2. Open the file "alice.txt" with encoding "latin-1". Print out all words of length more than five characters that appear more than five times in the file. Hint: Create an empty "counter" dictionary. Read the file line by line, generating a list of words for each line. Strip the word of punctuation symbols. If the resulting word is longer than five characters, insert the word into the counter dictionary if it not yet present, otherwise, if the word is present, then add one to the counter. Finally, walk through the dictionary printing out all words with more than five counts.
3. Open the file "mobydick.txt" with encoding "latin-1". For all words in the file longer than five characters that appear more than five times, print the word and the list of line-numbers in which they appear. Hint: Instead of having a dictionary that merely counts, we have a dictionary that takes as keys the words of length larger than five and as value a list of line numbers.
4. Use the archive HIV.txt, which contains the genome of various HIV-viruses. It is organized as a text files with various lines. The odd lines are short and contain information about the virus. The even lines are very long because they contain the actual genome encoding. Print out on separate lines
 - (a) the virus type
 - (b) the number of characters in the virus description
 - (c) the number of "A", "C", "G", and "T" characters for each virus
 - (d) the number of times two consecutive characters "AA", "AC", "AG", "AT", "CA", ... "TG" and "TT" appear in the genome.