

MSDS 7330

File Organization and Database Management

Homework Anagram

Name:

This is a homework assignment for MSDS 7330, File Organization and Database Management. Your final submission shall be a single pdf document that includes this document plus your answers to each of the questions.

Collaboration is expected and encouraged; however, each student must hand in their own homework assignment. To the greatest extent possible, answers should not be copied but, instead, should be written in your own words. Copying answers from anywhere is plagiarism, this includes copying text directly from the textbook. Do not copy answers. Always use your own words. For each question list all persons with whom you collaborated and list all resources used in arriving at your answer. Resources include but are not limited to the textbook used for this course, papers read on the topic, and Google search results. Note that 'Google' is not a resource. Don't forget to place your name on the document.

This assignment was adapted from the requirements of Dr. Eric Larson.

Python

In this assignment you will be using Python to investigate the scrabble dictionary. All the words in the scrabble dictionary are available from

<http://www.puzzlers.org/pub/wordlists/ospd.txt>

Using Python with Scrabble Word List

The text file contains a list of the words separated by newlines/carriage returns. Your first portion of the assignment is to find out how many unique anagrams are in the dictionary using Python. Note that this can be done efficiently in about 30 lines of code. However the method you use to store the words can wildly change the complexity of the problem. This can be completed procedurally or using object-oriented coding practices: either are fine implementations.

To find out what an anagram is, check out the Wikipedia page (<http://en.wikipedia.org/wiki/Anagram>). Essentially, anagrams are words with the same letters like 'ape' and 'pea'. This pair of words forms one anagram. As another example, 'ate', 'tea', and 'eat' form another anagram.

Turn in your Python code used to answer the following questions in addition to the answers to the questions.

Question 1 : How many unique anagrams are in the scrabble word list?

Collaborators:

Resources:

Question 2 : What is the anagram with the largest number of words in the word list? How many words are in this anagram?

Collaborators:

Resources:

Question 3 : Use matplotlib (or Excel) to visualize a histogram of the size of the anagrams. That is, make a bar chart where the X-axi from from $X = 1$ to $X = \text{max_anagram_size}$. Each bar will represent the number of anagrams of size X .

Collaborators:

Resources:

Using Python with Databases and Scrabble Word List

Create a database using Python with one table. The table will have three columns: Unique Anagram identifier (i.e. the sorted letters in the anagram), number of words in the anagram, and the actual words as a comma separated value string.

Use Python to create and populate the database. You can use MySQL to create a bare bones database, but Python should create the table and setup the variable types. Alternatively, you can use sqlite3 that ships with Python.

Now create queries to the database that answer the following questions (which are the same as Questions 1 and 2). Turn in your Python code used to answer these questions with the database.

Question 4 : How many unique anagrams are in the scrabble word list?

Collaborators:

Resources:

Question 5 : What is the anagram with the largest number of words in the word list? How many words are in this anagram?

Collaborators:

Resources:

Using Python with Databases and Merriam-Webster English Language Dictionary

Use Python to create a new table in the database and again find anagrams in the word list using three columns. However, instead of using the small dataset of scrabble words, use the Merriam-Webster English language dictionary (approximately 300,000 words).

Now create queries to the database that answer the following questions (which are the same as Questions 1 and 2). Turn in your Python code used to answer these questions with the database.

Question 6 : How many unique anagrams are in the Merriam-Webster word list?

Collaborators:

Resources:

Question 7 : What is the anagram with the largest number of words in the word list? How many words are in this anagram?

Collaborators:

Resources: