

COP-4338 System Programming

Programming Assignment 6: Data Structures in C

KF School of Computing & Information Sciences

In this assignment, you are asked to write a program that does the following:

- design and implement a hash table for storing strings in C (similar to what has been taught);
- reads the list of English words with correct spelling from the given file “dictionary.txt” and store it in the hash table;
- and finally, read the content of “input.txt” file and print out the words spelled incorrectly in the file.

1 Hash Table

For the sake of simplicity, please implement a hash table with 10,000 buckets that uses chaining to resolve collision. To design and implement a hash table, the most important question is how to select a good hash function. For this problem, you can use the following hash function which gets a string (char array) as its input parameter and returns a bucket index in the range of 0 to 9,999:

$$\text{hash}(\text{key}) = \left(\text{key}[n-1] + 17 * (\text{key}[n-2] + 17 * (\text{key}[n-3] + \dots + 17 * (\text{key}[1] + 17 * \text{key}[0])) \dots) \right) \% 10,000$$

where $n = \text{strlen}(\text{key})$ and $\text{key}[i]$ represents the ASCII code of the $(i+1)^{\text{th}}$ left-most letter of string key .

2 Spell-Checking a Given Corpus

In order to write a program that spell-checks the content of a given “input.txt” file, you need to first load the hash table up with the words listed in the “dictionary.txt” file. Then, you need to tokenize the content of “input.txt” file (assume that any character except *alphabetical characters*, *dash* and the *apostrophe character* is considered to be a delimiter that separates words). Finally, for each tokenized word, you must search the hash table to see if it exists

or not. If the search is not successful, your program must print out the word that has been misspelled.

3 Program Input

Assume that the name and address of both the dictionary.txt and input.txt files are given to the program as command-line arguments.

4 Program Output

Your program must print out all misspelled words on screen.

5 Submissions

You need to submit a *.zip* file compressing the C source file(s) related to the assignment (*.c* files) along with a README.txt file explaining the parts of assignment that you have successfully implemented.