CS242
System Programming Lab No. - 8

Objective
In this lab we learn about PERL.

Spell Checker

1. Write a PERL program that takes a sentence as input. For each word in the input sentence, the spell checker will do the following:
    *   test if the word exists in the dictionary. Use the dictionary /usr/share/dict/words
    *   If the word exists, then there is no spelling mistake and it will move on to the next word.
    *   If the word does not exist in the dictionary, then the word is misspelled. Print all the misspelled words.
2. Improve the spell checker program by comparing the misspelled word against words in the dictionary and return the most similar ones. Once all the misspelled words and their correct spellings are found, the spell checker will print them out to screen.

    a) Distance between words
    In order to find the distance between words, assume all *edits* have a cost. For example, the distance between "Hello" and "Help" is 2. One delete and one substitute. The steps to find distance are
    (i)   Initialization: Initialize the two strings in a two dimensional matrix

	H	**E**	**L**	**L**	**O**
0	1	2	3	4	5
H 1					
E 2					
L 3					
P 4					

    (ii)   Fill in each cell of the matrix using the following equation:
    $$m[i][j] = min (m[i][j-1] + 1, \ m[i-1][j] + 1, m[i-1][j-1] + cost)$$
    Cost is 0 if the *ith* character of the first string equals the *jth* character of the second string. *Cost* is 1 if they are not the same. The distance is in the bottom right cell.

	H	**E**	**L**	**L**	**O**
0	1	2	3	4	5
H 1	0				
E 2	1				
L 3	2				
P 4	3				

		H	E	L	L	O
	0	1	2	3	4	5
H	1	0	1	2	3	4
E	2	1	0	1	2	3
L	3	2	1	0	1	2
P	4	3	2	1	1	2

b) Efficiency

Comparisons and lookups can be expensive, so you should try to optimize the code. Hashes are one way to improve lookups. There are many other techniques which can be used.

c) Test Cases

You must also turn in a test case file for this assignment. This file will contain sample test cases and their expected outputs. The test case file should try to check all possible type of errors.

Deliverables

Create a pdf document containing the output of your program. Next create an archive file by the name <roll number>.tar that contains the pdf document and the C program. Mail the document to cs242@iitp.ac.in with subject "Lab <no>".

---000---