College of Computer & Information Science
Northeastern University
CS5800: Algorithms

Spring 2011
Handout 22
7 April 2011

Programming Assignment
Distance Between Words in a Dictionary
(Due Thursday, April 21, 5 PM)

The input to your program (to be read from the standard input) is:

1. An integer k, $1 < k < 20000$, on the first input line.
2. A set W of k distinct ASCII strings on the next k input lines. Each of these strings will have at most 25 characters.
3. A sequence of zero or more "commands" (following the set W). There are two valid kinds of commands: D and P. Each command occupies three input lines. The first of these three lines consists of a single character (D or P) followed by a newline. The second and third lines contain two distinct strings α and β drawn from W.

For each command, your program should perform the corresponding action as described below.

1. **D command**. Your program should report the "distance" between strings α and β. This value should be printed on a single output line, with no preceding blanks. The distance between α and β is the minimum number of single-character deletion/insertion operations needed to transform one string into the other.

   Examples: the distance between strings "Fred" and "frazzled" is 6 – delete 'F', insert 'f', insert 'a', insert 'z', insert 'z', insert 'l'. The distance between strings "leg" and "gel" is 4 – delete 'l', insert 'g', delete 'g', insert 'l'.)

   The D command is best solved using a dynamic programming algorithm. The procedure is similar to the edit distance problem we have considered in class. The exact definition of distance is different here, since we only allow deletion and insertion of characters.

2. **P command**. Let G be the undirected graph over vertex set W, in which there is an edge between two strings if and only if the distance between them is at most 2. If there is a path between the given strings α and β, then your program report the sequence of strings corresponding to the vertices of the path. The strings should be printed one per output line, beginning with string α and ending with string β. If there is no path between the two strings, then your program should report "No path".

Your program should terminate once it has processed all of the commands. You can assume that the input will be given in the proper format. (For example, you can assume that the two arguments of any command will be distinct strings drawn from W).

You may write the program in any mainstream programming language (C, C++, Java, Scheme, Python). Your submission must include full details on how to compile and execute your code. More information on the submission process will be provided soon.

The assignment will be graded out of 30 points, i.e., it is worth 3% of the overall total number of points.