

**22C : 031 (CS : 3310 : 0001) Algorithms**  
**Approximate String Matching – Homework**

1. Use a programming language you are familiar with (preferably Java) to implement the edit-distance algorithm you learned in class. Basically, you should define a function `editDistance(str1, str2)` taking two strings *str1* and *str2* as input, and returning the edit distance between them.
2. In this question, you apply the function `editDistance(str1, str2)` implemented in question 1 for a small application on spell correction. You are given a file called *dictionary.txt* containing a list of English words (one in each line), and a file called *misspelledWords.txt* containing a list of words misspelled by a user (See the illustration bellow). For each misspelled word, for instance, “*masachusets*”, your program needs to find the most similar word in file *dictionary.txt* (i.e. having the smallest edit distance), that is “*massachusetts*” in this example.

dictionary.txt

```
acknowledgement
accommodate
consensus
exaggerate
fundamental
massachusetts
perse
```

misspelledWord.txt

```
concensus
fundamental
masachusets
perveserance
```

3. What is the percentage of words for which your program returns the word that you would return?