Author: Ryan Timbrook
UW Net ID: timbrr
Project: Ling 473 Project 6 (Optional Extra Credit)
Date: September 11, 2016

Description:

This project is composed to two parts, the second building upon the first. The core component is to calculate the edit distance between two short texts. A text is defined as a sequence of sentences for this project. The second part of the project is to use the first components function to report the difference between the two texts which are being compared. The intent is measure the edit distance between two revisions of an article on the British Petroleum oil spill.

Approach:

The approach I took for this effort was to use python as the programming language and use dynamic programming to perform the edit distance calculations and alignment report print out. My edit distance function produces the expected validation edit distance value of the Lady Gaga comparison. However my alignment print out doesn't match exactly to what is shown in the Project 6 instructions. I believe I have an error in my 'alignText' function where I'm searching backwards through the distance matrix to find the lowest distance value of the three possible cells which represent possible back pointers.

I created an EditDistance class which is comprised of utility functions to process and maintain memory of the edit distance matrix. The inner line comparison is handled by the 'costSubstitution' function which based on a function identifier determines if a new instance of the EditDistance class should be created to handle the calculation of line distances. The final edit distance results of the inner function is returned to the calling instance of the EditDistance class and added to the text matrix.

Overall it's a challenging problem to align text and will dig into this topic more.