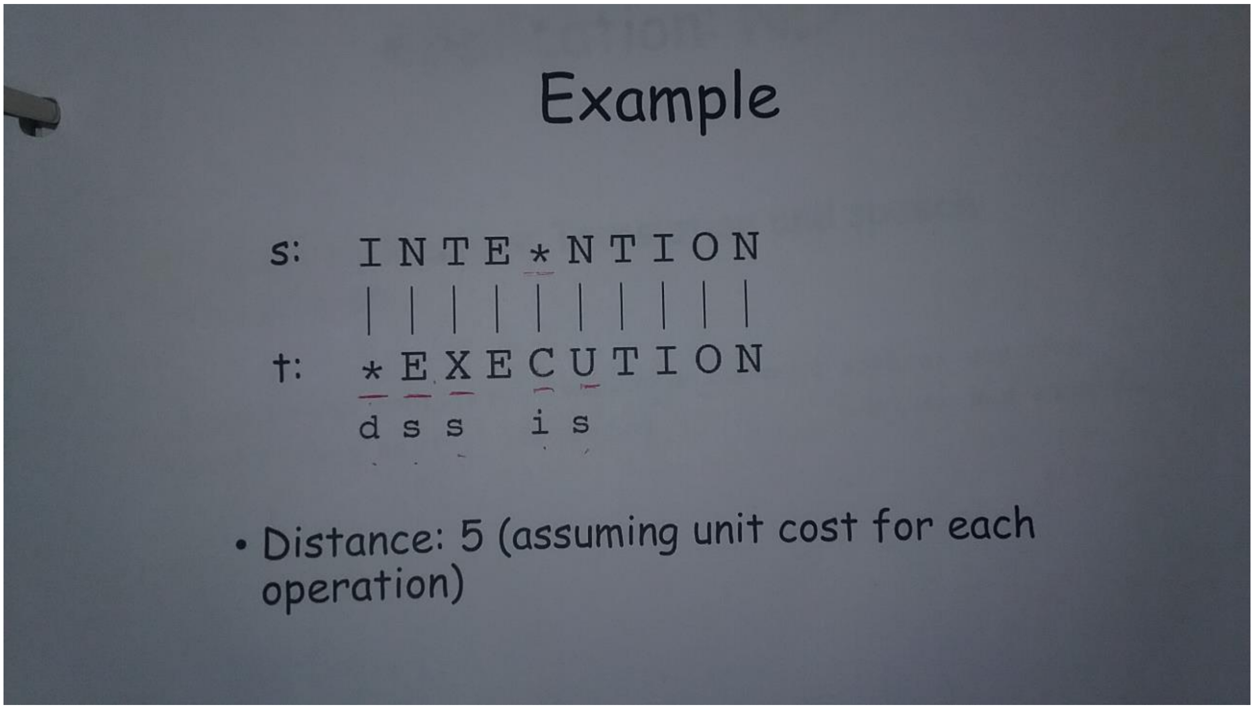
Week 4 Thurday Lecture Notes

Prep:

- Read DVP **Chapter 6.3**
- **Implementation 1** Due tonight @Midnight.
- *There is an update to the specifications*: No Longer need to output every single pair that is tied closest together.
- ****You only need to output the distance of the shortest and one of the pairs of points.**
- Might be useful to make input file changeable for Implementation 1.

Dynamic Programming: Edit Distance:

- Minimum number of edit operations needed to turn s into t. (arbitrary strings t and s)
- Editing Operations:
    - I = Insertion
    - D = Deletion
    - S = Substitution



- **d = deletion || s = substitution || I = insertion.**

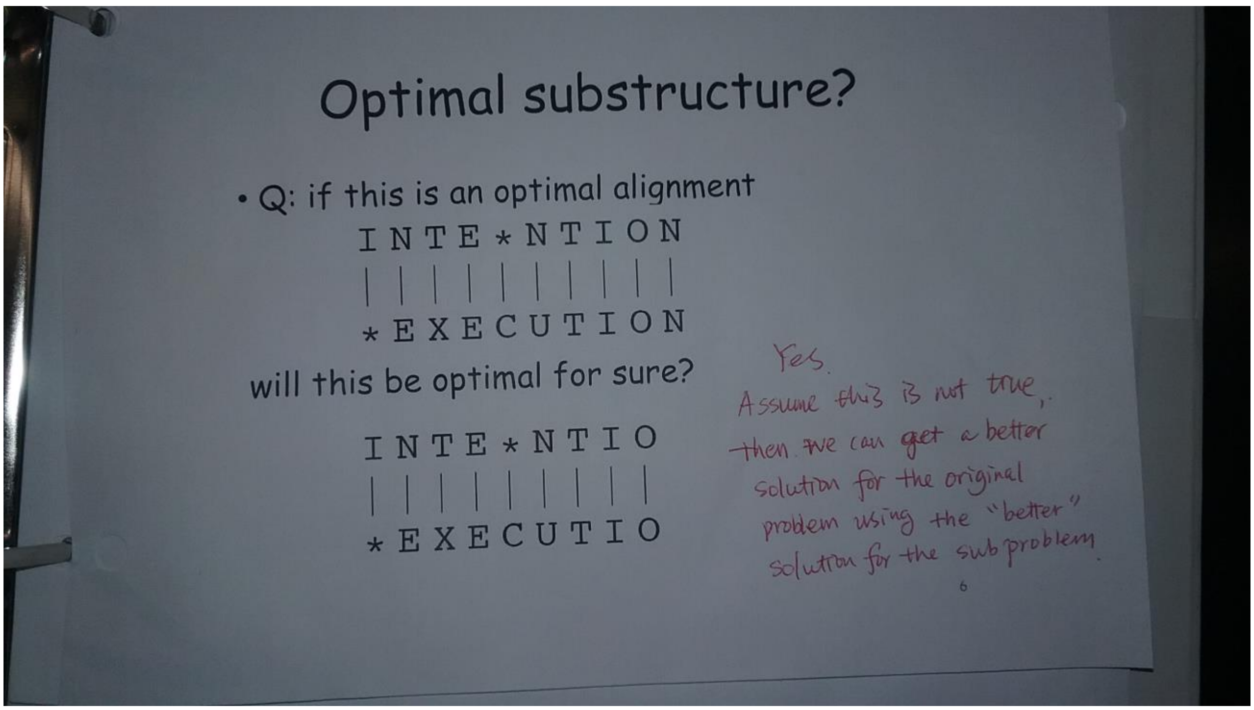
Finding the number of operations to change. String s to String t

Trying to find the most optimal edit distance.

Real Life Application:

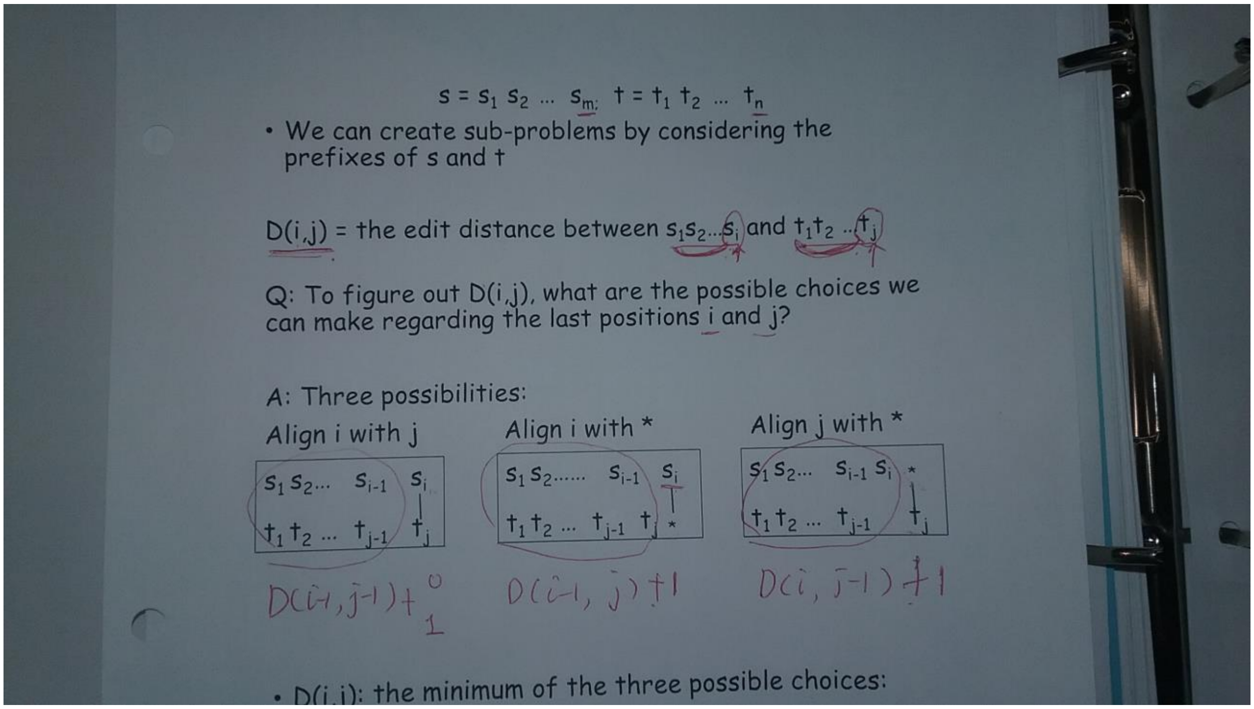
- Computational Biology
- Google Translation and speech recognition. (Very Often will not be a 1-1 Language Conversion)

Optimal substructure. If substructures have optimal alignment will them combined be the overall optimal structure? **YES**



Picture Shows that you can Insert, Delete, Or Align the Strings

D(i,j) = the edit distance between $s_1, s_2 \ldots s_i$ and $t_1, t_2 \ldots t_j$



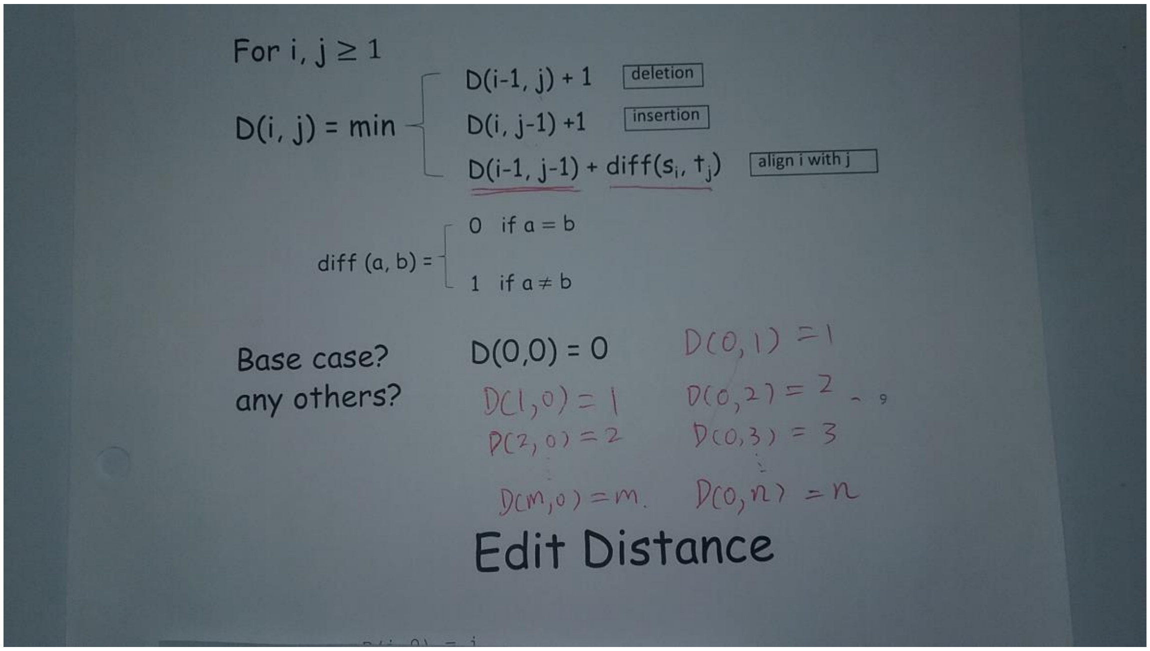
Picture Shows 3 Possibilities: Align, Insert, and Delete.

Align i with j: D(i-1, j-1) + 0 or 1 (0 for same, 1 for substitution)

Align i with *: D(i-1,j) + 1

Align j with *: D(i,j-1) +1

Recurrence Relation for D(i,j):



Picture Shows examples of why the base cases are true

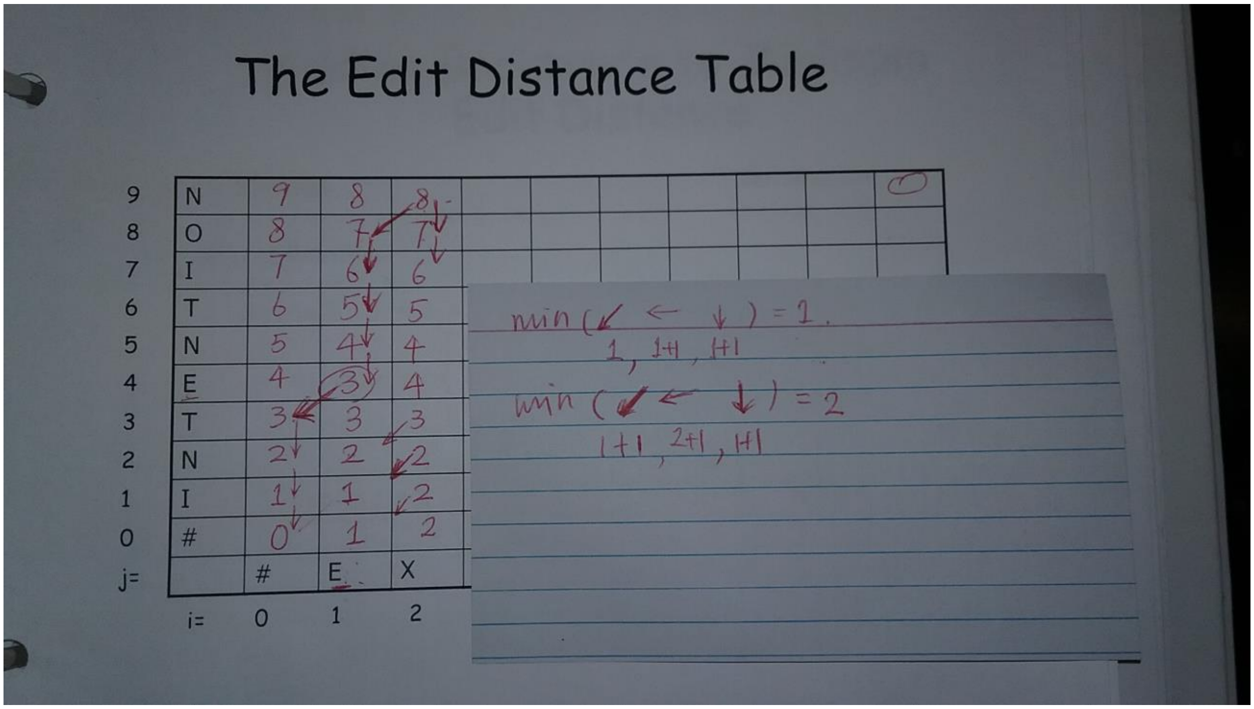
D(0,n) = n

D(0,0) = 0

D(n,0) = m

Edit Distance: Pseudocode. Runtime speed of O(m*n) roughly a quadratic function runtime (very expensive)

Edit Distance Table: How expensive it is for each position:



***Picture shows the Backtrace (always start in the top right node)**

For This Table example we start at point [2,9] = 8.

A possible Path is:

8,7,6,5,4,3,3,2,1,0.

There may be **multiple Optimal Paths.**

Computing Alignment

- We do this by keeping a "Backtrace"
- Getting the edit distance isn't sufficient. We often need to align each character of the two strings to each other.
- Trace back the path from the upper right corner to read off the alignment.
- See Table picture of **Traceback**
- Adding "backtrace" to minimum edit distance.

Preformance:

Time: O(nm)

Space: O(nm)

Backtrace: O(n+m)

Ending Notes:

- **Implementation Assignment 1:** Due Midnight 2/2/2017
- (**You can submit up to 24 hours late without penalty**)
- **Contact TA/Instructor** with questions about the implementation.
- **Set 3 Quiz Problems** will probably be available over the weekend to study.
- **Quiz 3** will be next Thursday
- (Recitation for Quiz Review at normal location **Wednesday 5:45-6:45pm**)

Finished Dynamic Programming Edit Distance Lecture (W4D2)

End of Week 4 Thursday Lecture Notes

~Information composed by Notetaker Scott Russell for CS 325 **DAS** student