2019-04-08 Levenshtein Edit Distance

Monday, April 8, 2019 3:03 PM

Question #3 on exam:

$T\left(\frac{n}{2}\right) + \log N$

$N^{\log_2(1+\epsilon)}$ } is this not valid or case #3?

Levenshtein Edit Distance

- Goal of an edit distance algorithm is to determine how many transformations are needed in order to turn one word into another.
- To make these transformation, we have access to three operations
  - Insert new letter
  - Remove a letter
  - Substitute a letter

Example: Dog and Frog

- How many transformations are required to convert dog into frog?
- Possibility #1:
  - Add "F" at 0
  - Transform "D" into "R"

DOG → FDOG → FROG
#1 #2

Example #2: Translate Raccoon into Racket

- Possibility #1:

RACCOON → RACKOON → RACKEON
RACKETN ← RACKETN ↓

- Possibility #2:

RACCOON → RACKOON → RACKON → RACKEN
↓
RACKET

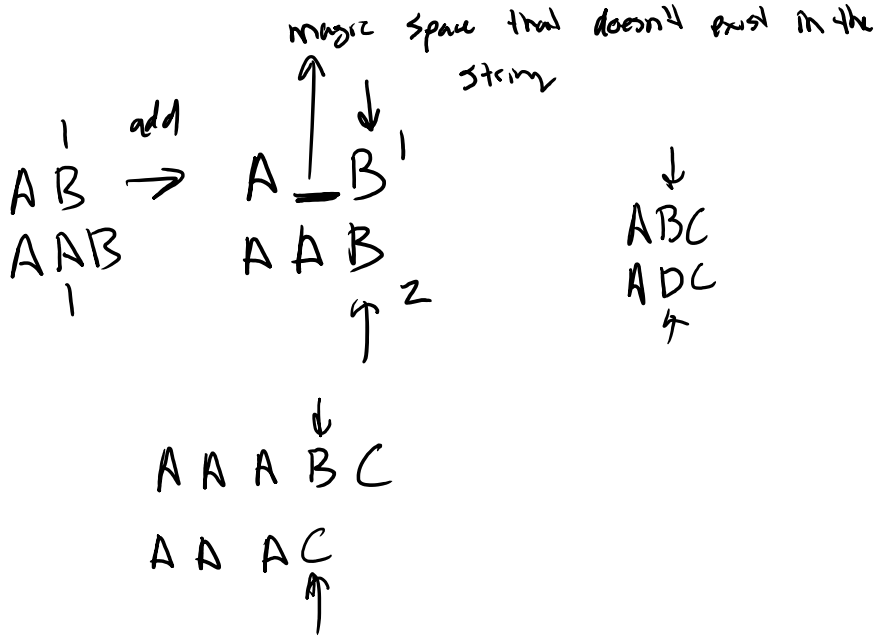
RACKET

AAABBC

AAC

If we do transformations and drop the rest, we get 4 ops but if we delete the middle elements, we get 3 ops

- Start by simultaneously examining each index starting at 0
- If they line up perfectly (e.g. s1[0] == a and s2[0] == a), nothing needs to be done
- Otherwise, the character at index is a candidate for:
  - Insert
  - Remove
  - Transform / replace
- The correct choice is determined by the final answer (lowest # of edits)



	F	R	O	G
D	2	1	1	2
O	2	1	1	1
G	3	2	1	1