**CS 5316 – Natural Language Processing**

**Quiz 2 Solution**

**(Time limit: 12 minutes)**

Instructions: (1) Please write legibly. Unreadable answers will NOT be graded; (2) Write in the spaces provided for the questions only.

1. (5 points) Find the minimum edit distance (Levenshtein) and show the corresponding alignment between GRAPHS and RAFTS. Use the following grid to show your computation..

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | # | R | A | F | T | S |  |  |
| S | 6 | 5 | 4 | 5 | 6 | 5 |  |  |
| H | 5 | 4 | 3 | 4 | 5 | 6 |  |  |
| P | 4 | 3 | 2 | 3 | 4 | 5 |  |  |
| A | 3 | 2 | 1 | 2 | 3 | 4 |  |  |
| R | 2 | 1 | 2 | 3 | 4 | 5 |  |  |
| G | 1 | 2 | 3 | 4 | 5 | 6 |  |  |
| # | 0 | 1 | 2 | 3 | 4 | 5 |  |  |
|  | # | R | A | F | T | S |  |  |

G R A P H S

\* R A F T S

1. (1 point) Compute the normalized longest common subsequence between GRAPHS and RAFTS.

= lcs(graphs, rafts)/[minlength(graphs, rafts)+edist(graphs, rafts)]

= 2 / (5+5) = 1/5

1. (4 points) Given the following running text, estimate P(sun | the ) using (a) MLE, and (b) add-one smoothing. Lemmatize before defining the vocabulary. .

The sun rises from the east, traverses the sky over the southern horizon, and sets in the west.

V = { the sun rise from east traverse sky over southn horizon and set in west }

|V| = 14

MLE: 1/5

Add-one: (1+1)/(5+14) = 2/19