West Visayas State University

COLLEGE OF INFORMATION AND COMMUNICATIONS TECHNOLOGY
Luna St., La Paz, Iloilo City 5000
Iloilo, Philippines
* Trunkline: (063) (033) 320-0870 loc 1403 * Telefax No.: (033) 320-0879
* Website: www.wvsu.edu.ph * Email Address: cict@wvsu.edu.ph

Assignment for Unit 3.2

Name: Jephone Israel Jireh S. TORRE Date: 03 – 08– 25

Year and Section: BSCS 3A AI

Note: This is not a coding activity. Put your answers directly in the answer sheet.

Given the following corpora, do the following:

Photosynthesis is the process by which plants convert sunlight into energy. During photosynthesis, chlorophyll in plant cells absorbs light and converts it into chemical energy. This energy is used to transform carbon dioxide and water into glucose and oxygen. The oxygen is released into the atmosphere, supporting life on Earth.

1. (20 points) Create a Bigram Language Model

   Bigram Count:

   ('energy', '.'): 2, ('photosynthesis', 'is'): 1, ('is', 'the'): 1, ('the', 'process'): 1, ('process', 'by'): 1, ('by', 'which'): 1, ('which', 'plants'): 1, ('plants', 'c 1, ('plants', 'convert'): 1, ('convert', 'sunlight'): 1, ('sunlight', 'into'): 1, ('into', 'energy'): 1, ('.', 'during'): 1, ('during', 'photosynthesis'): 1, ('photosynthesis', ','): 1, (',', 'chlorophyll'): 1, ('chlorophyll', 'in'): 1, ('in', 'plant'): 1, ('plant', 'cells'): 1, ('cells', 'absorbs'): 1, ('absorbs', 'light'): 1, ('light', 'and'): 1, ('and', 'converts'): 1, ('converts', 'it'): 1, ('it', 'into'): 1, ('into', 'chemical'): 1, ('chemical', 'energy'): 1, ('.', 'this'): 1, ('this', 'energy'): 1, ('energy', 'is'): 1, ('is', 'used'): 1, ('used', 'to'): 1, ('to', 'transform'): 1, ('transform', 'carbon'): 1, ('carbon', 'dioxide'): 1, ('dioxide', 'and'): 1, ('and', 'water'): 1, ('water', 'into'): 1, ('into', 'glucose'): 1, ('glucose',

'and'): 1, ('and', 'oxygen'): 1, ('oxygen', '.'): 1, ('.', 'the'): 1, ('the', 'oxygen'): 1, ('oxygen', 'is'): 1, ('is', 'released'): 1, ('released', 'into'): 1, ('into', 'the'): 1, ('the', 'atmosphere'): 1, ('atmosphere', ','): 1, (',', 'supporting'): 1, ('supporting', 'life'): 1, ('life', 'on'): 1, ('on', 'earth'): 1, ('earth', '.'): 1

2. (60 points) Calculate the Perplexity of the Bigram model created using the test sentence, use Laplace Smoothing to deal bigrams with 0 occurrences:

   "During photosynthesis, chlorophyll is activated by visible light, and an electron is promoted from its ground state to an excited state."

   Show your solution on how you derived to the score.

   Calculating Log Probability Sum:

   P(photosynthesis | during) = (1 + 1) / (1 + 1 * 39) = 0.05
   log2(0.05) = -4.321928094887363

   P(, | photosynthesis) = (1 + 1) / (2 + 1 * 39) = 0.04878048780487805
   log2(0.04878048780487805) = -4.357552004618084

   P(chlorophyll | ,) = (1 + 1) / (2 + 1 * 39) = 0.04878048780487805
   log2(0.04878048780487805) = -4.357552004618084

   P(is | chlorophyll) = (0 + 1) / (1 + 1 * 39) = 0.025
   log2(0.025) = -5.321928094887363

   P(activated | is) = (0 + 1) / (3 + 1 * 39) = 0.023809523809523808
   log2(0.023809523809523808) = -5.392317422778761

   P(by | activated) = (0 + 1) / (0 + 1 * 39) = 0.02564102564102564
   log2(0.02564102564102564) = -5.285402218862249

P(visible | by) = (0 + 1) / (1 + 1 * 39) = 0.025
log2(0.025) = -5.321928094887363

P(light | visible) = (0 + 1) / (0 + 1 * 39) = 0.02564102564102564
log2(0.02564102564102564) = -5.285402218862249

P(, | light) = (0 + 1) / (1 + 1 * 39) = 0.025
log2(0.025) = -5.321928094887363

P(and | ,) = (0 + 1) / (2 + 1 * 39) = 0.024390243902439025
log2(0.024390243902439025) = -5.357552004618084

P(an | and) = (0 + 1) / (3 + 1 * 39) = 0.023809523809523808
log2(0.023809523809523808) = -5.392317422778761

P(electron | an) = (0 + 1) / (0 + 1 * 39) = 0.02564102564102564
log2(0.02564102564102564) = -5.285402218862249

P(is | electron) = (0 + 1) / (0 + 1 * 39) = 0.02564102564102564
log2(0.02564102564102564) = -5.285402218862249

P(promoted | is) = (0 + 1) / (3 + 1 * 39) = 0.023809523809523808
log2(0.023809523809523808) = -5.392317422778761

P(from | promoted) = (0 + 1) / (0 + 1 * 39) = 0.02564102564102564
log2(0.02564102564102564) = -5.285402218862249

P(its | from) = (0 + 1) / (0 + 1 * 39) = 0.02564102564102564
log2(0.02564102564102564) = -5.285402218862249

P(ground | its) = (0 + 1) / (0 + 1 * 39) = 0.02564102564102564
log2(0.02564102564102564) = -5.285402218862249

P(state | ground) = (0 + 1) / (0 + 1 * 39) = 0.02564102564102564
log2(0.02564102564102564) = -5.285402218862249

P(to | state) = (0 + 1) / (0 + 1 * 39) = 0.02564102564102564
log2(0.02564102564102564) = -5.285402218862249

P(an | to) = (0 + 1) / (1 + 1 * 39) = 0.025
log2(0.025) = -5.321928094887363

P(excited | an) = (0 + 1) / (0 + 1 * 39) = 0.02564102564102564
log2(0.02564102564102564) = -5.285402218862249

P(state | excited) = (0 + 1) / (0 + 1 * 39) = 0.02564102564102564
log2(0.02564102564102564) = -5.285402218862249

P(. | state) = (0 + 1) / (0 + 1 * 39) = 0.02564102564102564
log2(0.02564102564102564) = -5.285402218862249

Log Probability Sum: -119.28407538297431

Perplexity = $2^{-\frac{Log\ Probability\ Sum}{N}}$

Perplexity = $2^{-\frac{(-119.28407538297431)}{23}} = 2^{5.1862650218641} = 36.410033096862946$

A perplexity of 36.410033096862946 indicates that the model has some predictive ability, but there is room for improvement.

West Visayas State University

COLLEGE OF INFORMATION AND COMMUNICATIONS TECHNOLOGY
Luna St., La Paz, Iloilo City 5000
Iloilo, Philippines
* Trunkline: (063) (033) 320-0870 loc 1403 * Telefax No.: (033) 320-0879
* Website: www.wvsu.edu.ph * Email Address: cict@wvsu.edu.ph

3. (20 points) Look the bigram probability calculation attached on my Bigram implementation: https://github.com/JephoneTorre/Assignment-for-Unit-3.2.git

   Modify the method definition so that Laplace smoothing is applied on the probability calculation.