West Visayas State University

COLLEGE OF INFORMATION AND COMMUNICATIONS TECHNOLOGY

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

Exercise for Unit 3

Name:John Manuel Carado Date: March 9, 2025

Year and Section: BSCS 3-A

1. (10 points) Use the Wikipedia python module and access any topic, as you will use that as your corpus, with a word limit of 1000 words.

Example usage:

```
import wikipedia

# Fetch a Wikipedia page
page = wikipedia.page("Natural_language_processing")
text = page.content[:1000] # Limit to 1000 characters
```

- 2. (60 points) Train 2 models: a Bigram and Trigram Language Model, use the shared code as reference for bigram modeling, and update it to support trigrams.
- 3. (30 points) Using a test sentence "The quick brown fox jumps over the lazy dog near the bank of the river." OR generate your own test sentence, create a function that will determine the perplexity score for each trained model.
 - a. Bigram model perplexity -> Test Sentence "" -> Score:
 - b. Trigram model perplexity -> Test Sentence "" -> Score:

Upload your output in a public Github Repository named CCS249-EXERCISE-3-LASTNAME and attach the Github link here:

https://github.com/hydraadra112/CCS249-EXERCISE-3-CARADO.git