**Assignment1 Report**

Name:洪振庭

Institution (school):NTHU

Student ID:110000223

Platform (Colab/Kaggle/Local):Local

Python version:3.13.2

Operating system: Windows

CPU: 11th gen intel(r) core(tm) i5-11300H @ 3.10ghz

GPU requirement: nvidia geforce gtx 1650

1. Which embedding model do you use? What are the pre-processing steps? What are the hyperparameter settings? (5%)

Answer:

1. What will the performance be like if you sample 5%, 10% and 20% of wiki text in TODO4? (10%, 3% for each)

Answer:

1. What is the performance for different categories or sub-categories when trained on different corpora? (15%)

3.1 Present your results. (5%)

Answer:

3.2 Introduce the corpus you selected and explain the differences between the Wikipedia corpus and your corpus. (including data size, topic difference, structural difference … ) (5%)

Answer:

3.3 Explain why the accuracy increases or decreases. (5%)

Answer:

1. Select a few words and use their embeddings to retrieve the five most similar words and present the results. What do you observe? (10%)

Answer:

1. Anything that can strengthen your report. (5%)

Answer: