Name: Yue Ka Leung	Student ID: 1155214424	Major: Computer Science
Name: Tang Yu Hin	Student ID: 1155211754	Major: AIST

Project subject: A peek into word embeddings using word2vec and its related applications.

Description:
In this project, we will learn the principle of word2vec and explore the use of probability in calculating the similarity of words in word prediction using the Skip-Gram model. To achieve this goal, we plan to read related materials like research papers and online lectures, and eventually write a simulation in Python using pre-trained models after understanding the basics.

Planned activities (readings and simulation):

1. Research paper: https://github.com/yueagar/Word2Vec-bias-extraction
2. Geeks4Geeks: Word Embedding using Word2Vec - GeeksforGeeks
3. Stanford CS224N: Stanford CS224N: NLP with Deep Learning | Winter 2021 | Lecture 1 - Intro & Word Vectors (youtube.com)
4. Medium: A math-first explanation of Word2Vec | by Ankur Tomar | Analytics Vidhya | Medium
5. Our simulation: (we will create a new GitHub repository later and include the demo and the link in our presentation in the future)

Presentation outline:

1. Introduction to word2vec
2. Probabilistic Foundations of the Skip-Gram Model
3. Code implementation of word2vec
4. Findings and conclusion

Optional/possible activities (maybe too challenging):

1. Optimization of models: negative sampling – obtain the gradient vector (multivariable calculus). Negative sampling: https://mccormickml.com/2017/01/11/word2vec-tutorial-part-2-negative-sampling/
2. Analysis real-world articles to verify the probability theory and identify potential flaws and biases during word analysis (refer to Planned activities #1).