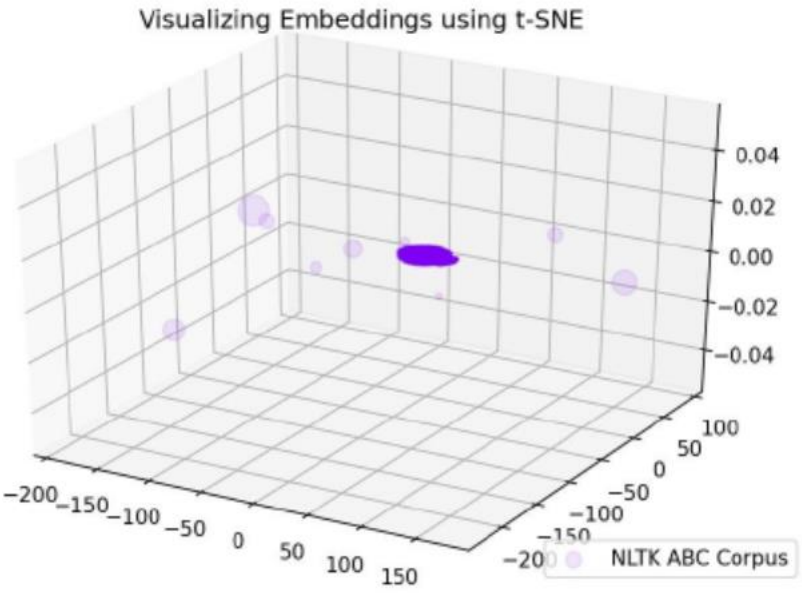
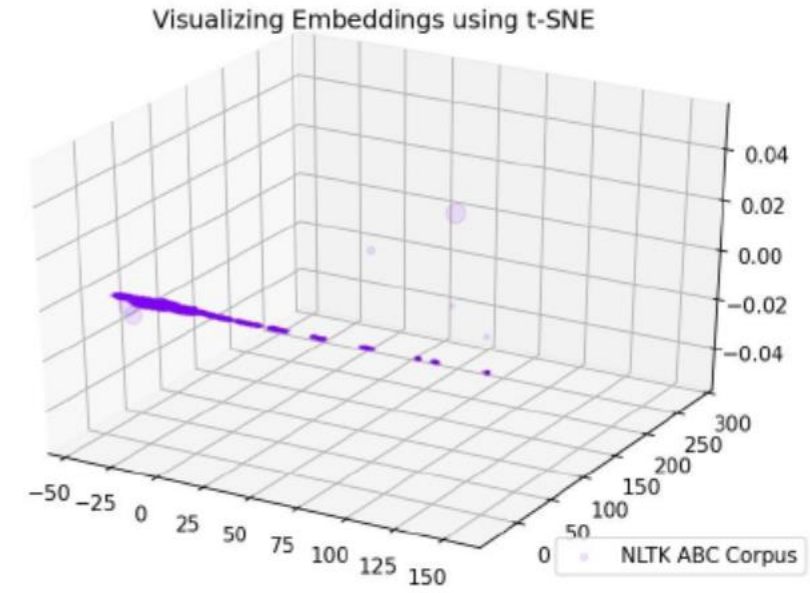
MCA Assignment 3

1.Word2Vec is an algorithm that allows to categorise words as vectors. It generates word embeddings and map vectors in vector space. It basically trains the words against the other neighbouring words in the corpus. The words which are used in similar context are given similar embeddings. It uses one hot vector (one bit 1 and others 0). In CBOW (continuous bag of words), we predict the target word using context. The centre word of window is predicted. One hot networks of neighbouring words are passed and the central word is predicted. Error is used to adjust the weights. Keep sliding the window and continue process. This algorithm is Skip-gram which takes input word and estimate probability of other words appearing close to that word.

During the training process, the words similar to each other start getting placed close to each other.

Average Loss: 0.31





2. Baseline Retrieval

MAP: 0.5183859040856561

Retrieval with Relevance Feedback

MAP: 0.5188705050655281

Retrieval with Relevance Feedback and query expansion

MAP: 0.6253043920389884

With relevance feedback and vector adjustment, MAP slightly increases. It might increase more if hyperparameters are tuned accordingly.

With relevance feedback and query expansion, MAP shows considerable increase. This result is in line with the expectation.