Mathematics of Natural Language Processing algorithms (cont'd)

Lab Exercise 4:

1. Calculate the bag of words (without TF-IDF, just word counts) matrix for the following 5 phrases.

- college student

- airport taxi driver

- Toronto art museum

- life art

- machine learning student

Hints: How many rows should be there? How many columns should be there?

2. Calculate the IDF for the words "college" and "art".

Let's switch to jupyter and open "NLP.ipynb" to study this in more detail.