

## Assignment 4 - Naive Machine Translation and LSH

You will now implement your first machine translation system and then you will see how locality sensitive hashing works. Let's get started by importing the required functions!

If you are running this notebook in your local computer, don't forget to download the twitter samples and stopwords from nltk.

```
nltk.download('stopwords')
nltk.download('twitter_samples')
```

**NOTE:** The `Exercise xx` numbers in this assignment **are inconsistent** with the `UNQ_Cx` numbers.

### This assignment covers the following topics:

- [1. The word embeddings data for English and French words](#)
  - [1.1 Generate embedding and transform matrices](#)
    - [Exercise 1](#)
- [2. Translations](#)
  - [2.1 Translation as linear transformation of embeddings](#)
    - [Exercise 2](#)
    - [Exercise 3](#)
    - [Exercise 4](#)
  - [2.2 Testing the translation](#)
    - [Exercise 5](#)
    - [Exercise 6](#)
- [3. LSH and document search](#)
  - [3.1 Getting the document embeddings](#)
    - [Exercise 7](#)
    - [Exercise 8](#)
  - [3.2 Looking up the tweets](#)
  - [3.3 Finding the most similar tweets with LSH](#)
  - [3.4 Getting the hash number for a vector](#)
    - [Exercise 9](#)
  - [3.5 Creating a hash table](#)
    - [Exercise 10](#)
  - [3.6 Creating all hash tables](#)
    - [Exercise 11](#)

```
In [1]: import pdb
```

