

Lab Exercise 10 - Pre-trained Models

In this lab exercise, we will **create document vectors** from pre-trained models.

1. Download the **pre-trained** models for the following word embedding models - check the notebooks uploaded.
 - a. GloVe
 - b. Word2Vec
2. Create **document vectors** by the following formula:

$$\text{doc_Vec}_i = \sum_j w_j \quad \text{where:}$$

- a. **doc_Vec_i** : i^{th} document in the corpus.
- b. **w_j** : word vector of j^{th} word in the document. The word vector is taken model.

For example for a document $d = \text{"I like apple"}$. Its doc_Vec is created as follows:

doc_Vec	I	like	apple
1.696285	0.673522	0.033174	0.989588
0.803228	-0.34009	0.162396	0.98092
-0.76719	0.110601	-0.41541	-0.46238
-0.19525	-0.95803	0.361487	0.401287
1.78991	0.872004	0.650369	0.267536
0.967644	0.227021	0.989982	-0.24936

- c. (Out-Of-Vocabulary) **OOV** words can be **ignored**.
3. (*Optional*) For fastText, download a **non-English** language. Test few words similarities.