☰ **Hide menu**

How to Create Word Embeddings?

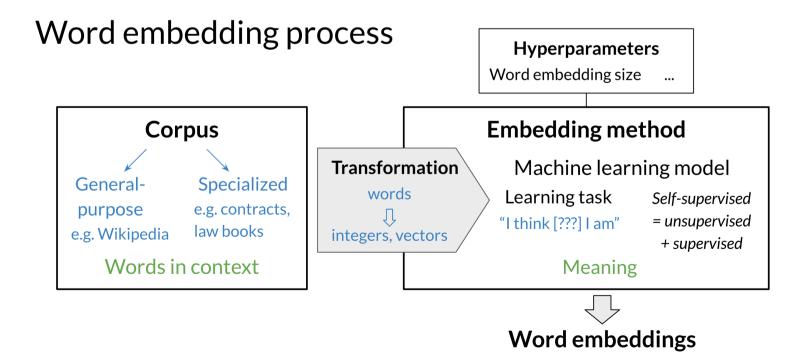To create word embeddings you always need a corpus of text, and an embedding method.

The context of a word tells you what type of words tend to occur near that specific word. The context is important as this is what will give meaning to each word embedding.

Embeddings

There are many types of possible methods that allow you to *learn* the word embeddings. The machine learning model performs a learning task, and the main by-products of this task are the word embeddings. The task could be to learn to predict a word based on the surrounding words in a sentence of the corpus, as in the case of the continuous bag-of-words.

The task is **self-supervised**: it is both unsupervised in the sense that the input data — the corpus — is unlabelled, and supervised in the sense that the data itself provides the necessary context which would ordinarily make up the labels.

When training word vectors, there are some parameters you need to tune. (i.e. the dimension of the word vector)