1. Vector-Space Model
2. Create a dictionary of words present in the documents (exclude stop words, only extract important features);
   1. Select all terms;
   2. Convert it to a dimension in a vector space;
3. Build feature vectors for each documents using word frequency;
4. Tokenization, Counting and Normalization is called the **Bag of Words;**