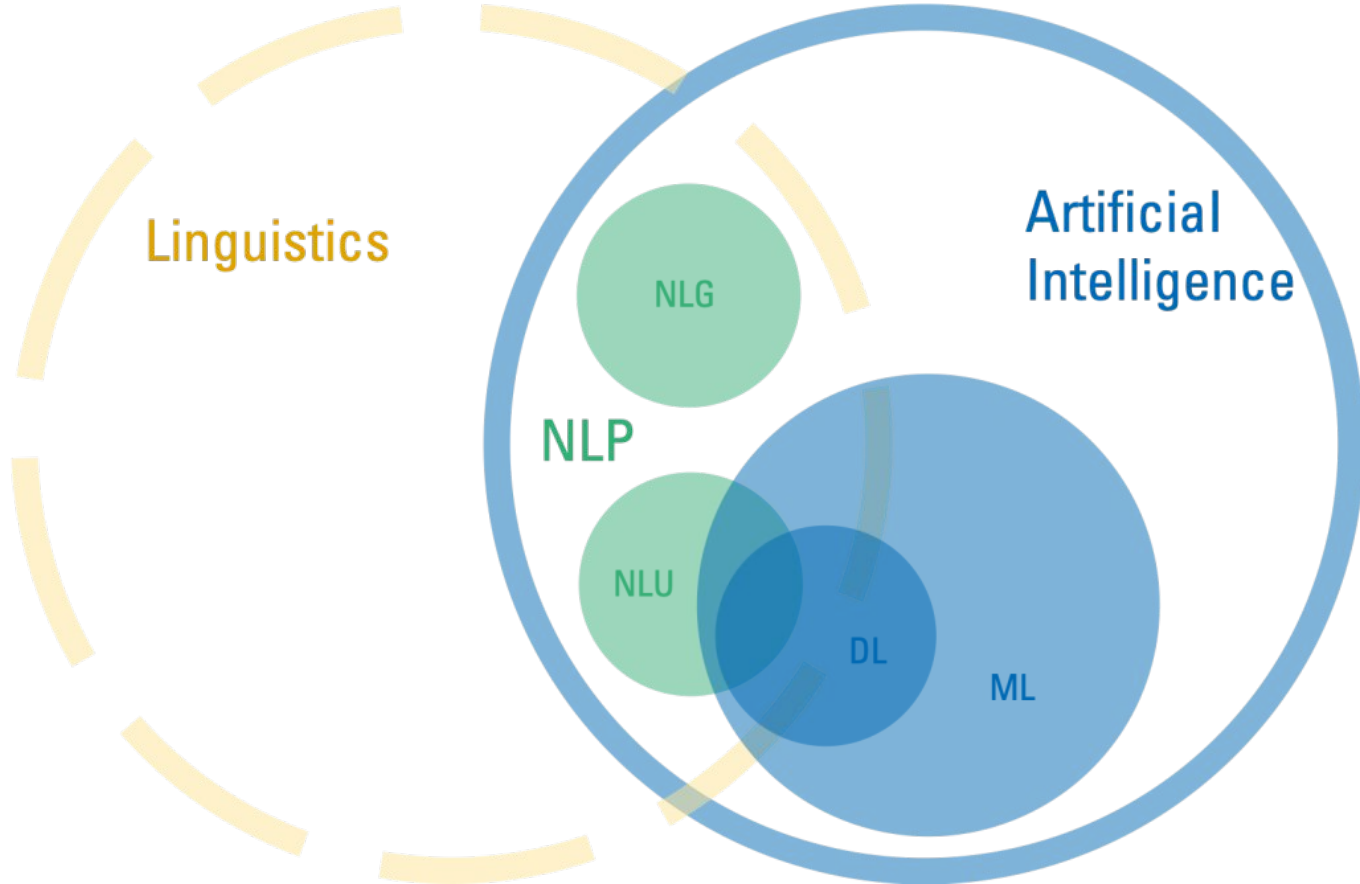


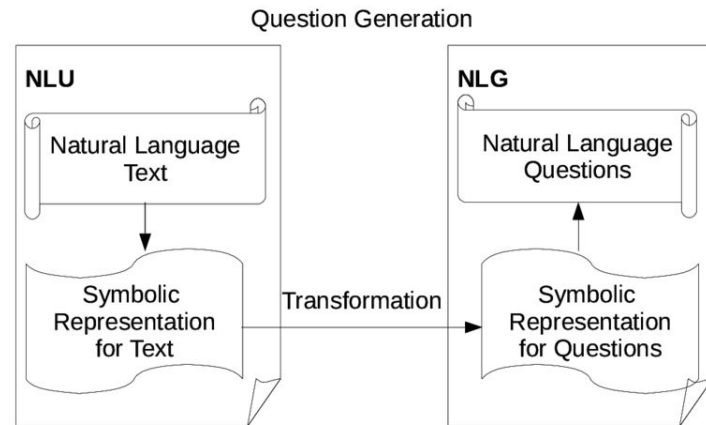
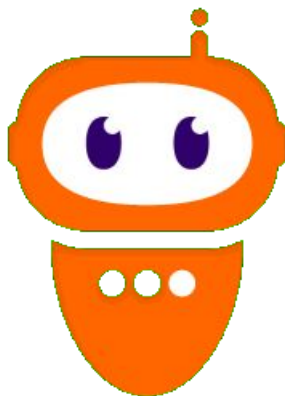
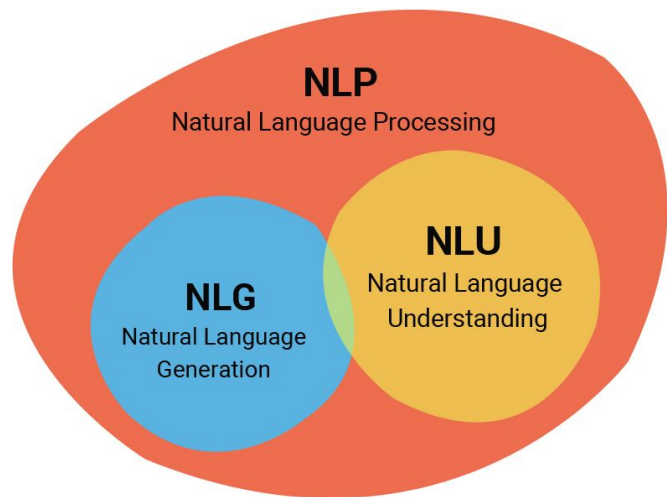
INTRODUCTION TO NATURAL LANGUAGE PROCESSING

DR.CHRISTELLE SCHARFF

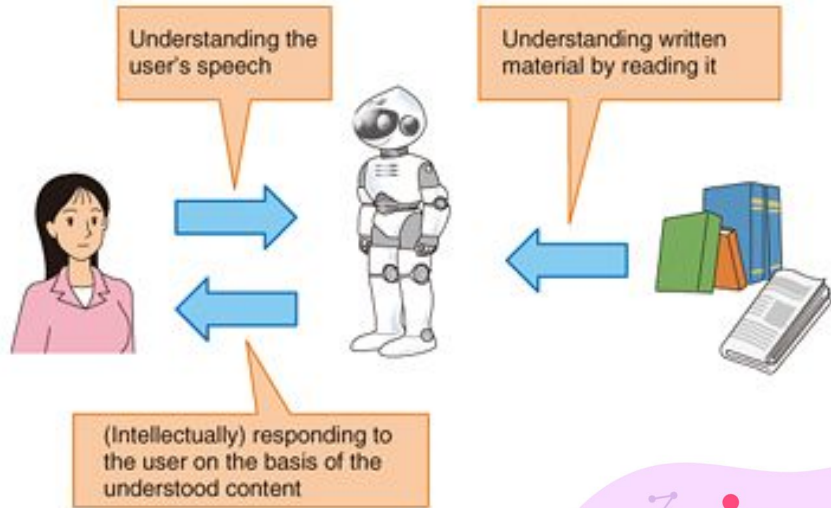
NATURAL LANGUAGE PROCESSING (NLP)



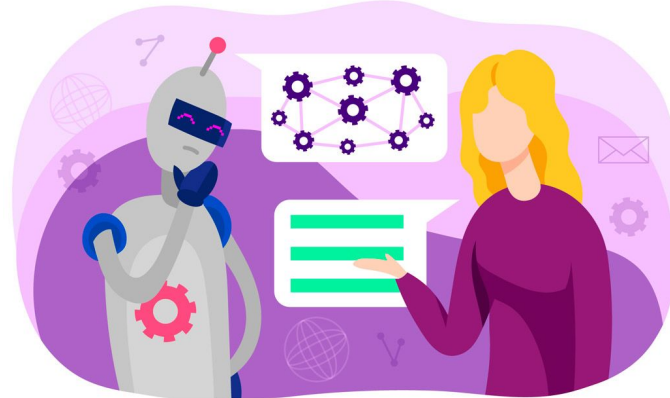
MORE ABOUT NLP



NLP - APPLICATIONS



Natural Language Processing



NLP TERMINOLOGY

NER

CORPORA

SENTIMENT ANALYSIS

BAG OF WORDS



NLP TERMINOLOGY

**WORD SENSE
DISAMBIGUATION**

ESA

LDA

**LATENT SENTIMENT
ANALYSIS (LSA)**



TOKENIZATION

Natural Language Processing
['Natural', 'Language', 'Processing']

Address Q&A

What is your address ?

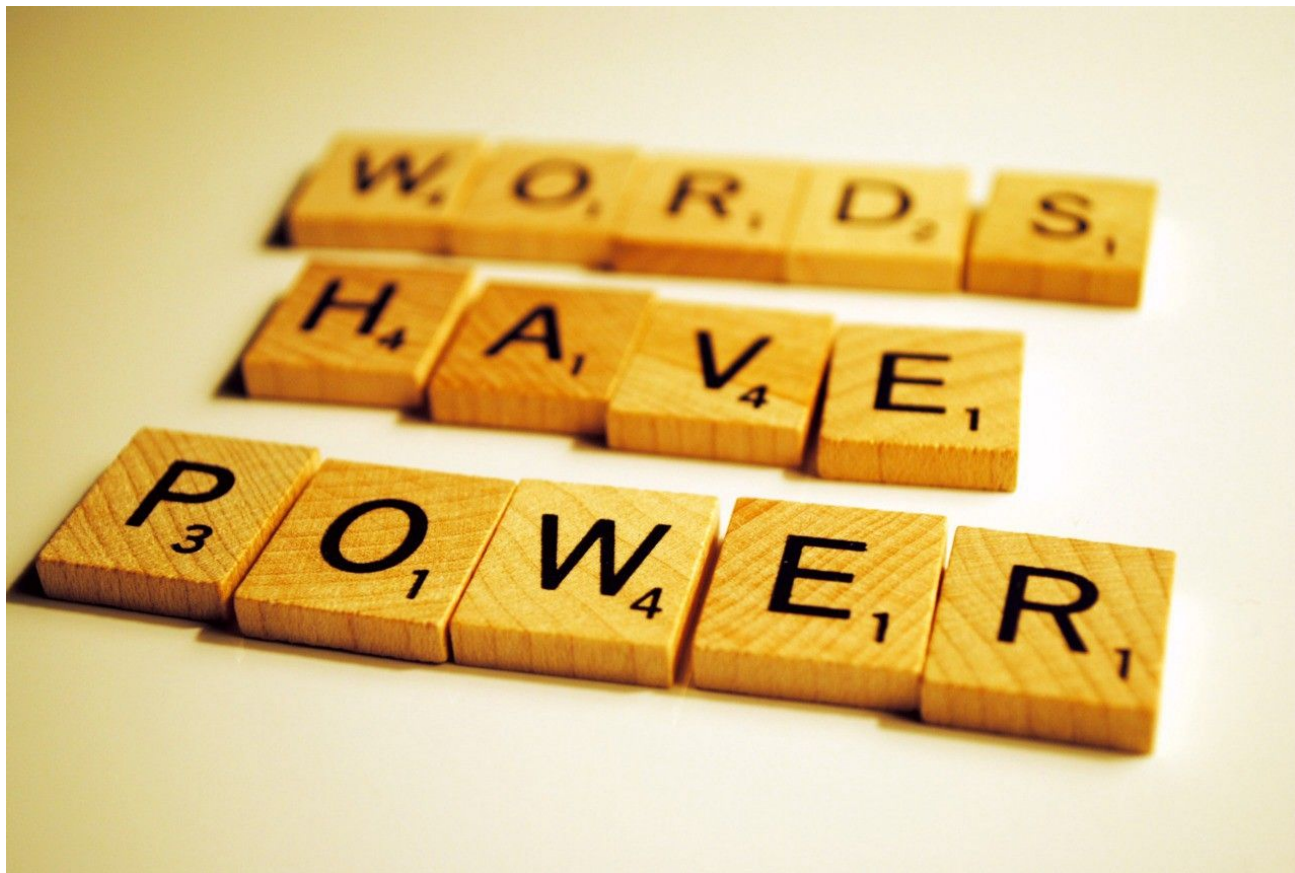
Where are you located ?

Phone Q&A

What phone can I reach you on ?

What is your phone number ?

BAG OF WORDS

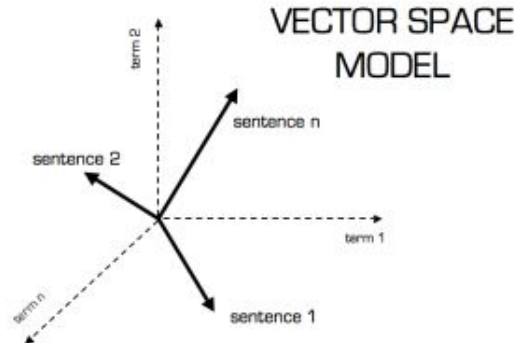


EXAMPLE OF BOW IN MOVIE REVIEWS

	good	movie	not	a	did	like
good movie	1	1	0	0	0	0
not a good movie	1	1	1	1	0	0
did not like	0	0	1	0	1	1

VECTORIZATION

Word Vector Representations



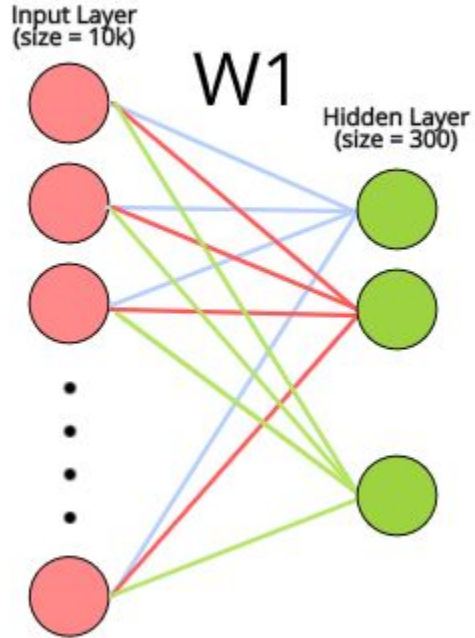
VECTORIZATION

$$TF(t) = \frac{\text{Number of times term } t \text{ appears in a document}}{\text{Total number of terms in the document}}$$

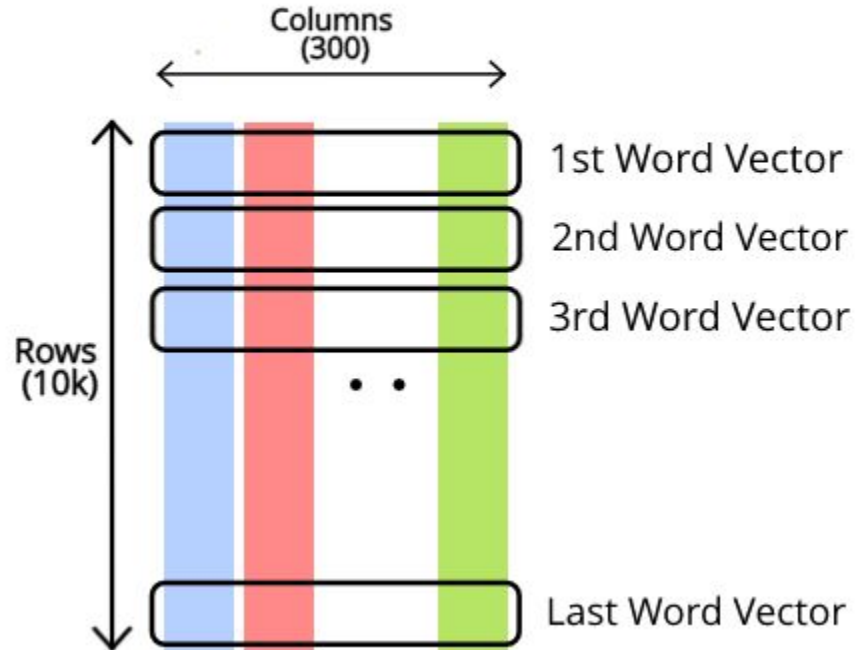
$$IDF(t) = \log_e \left(\frac{\text{Total number of documents}}{\text{Number of documents with term } t \text{ in it}} \right)$$

$$TF - IDF \text{ score} = TF * IDF$$

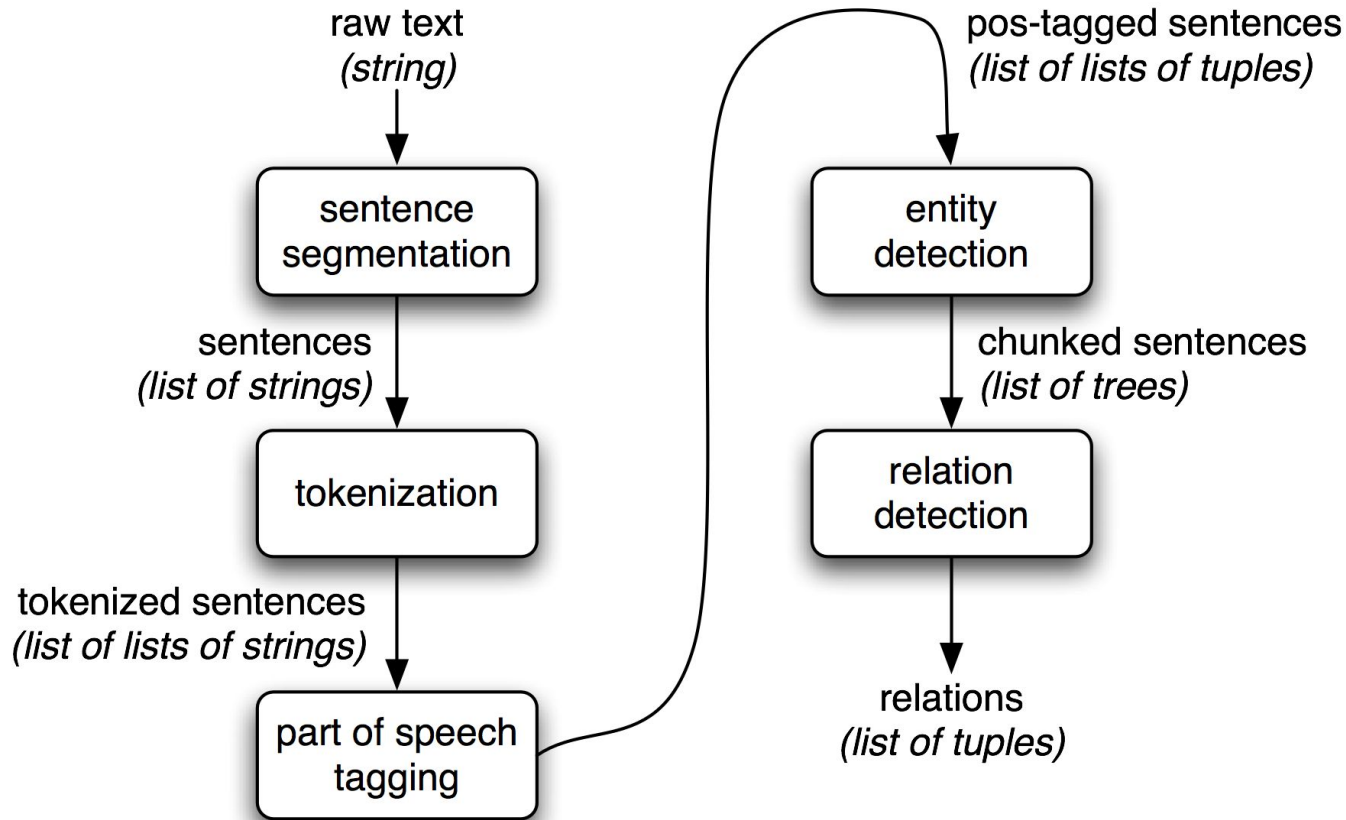
WORD2VEC



Weight Matrix



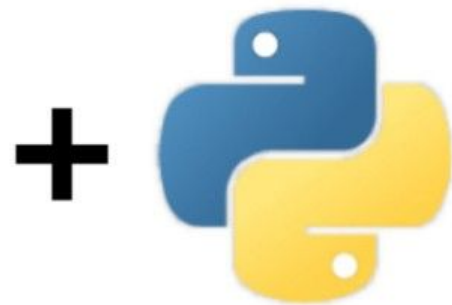
NLP INFORMATION EXTRACTION ARCHITECTURE



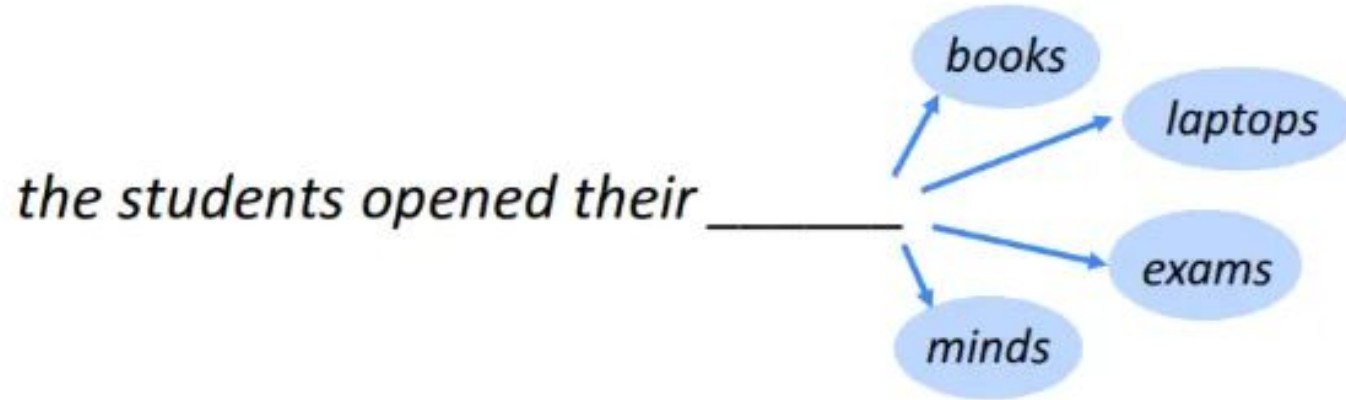
NLTK



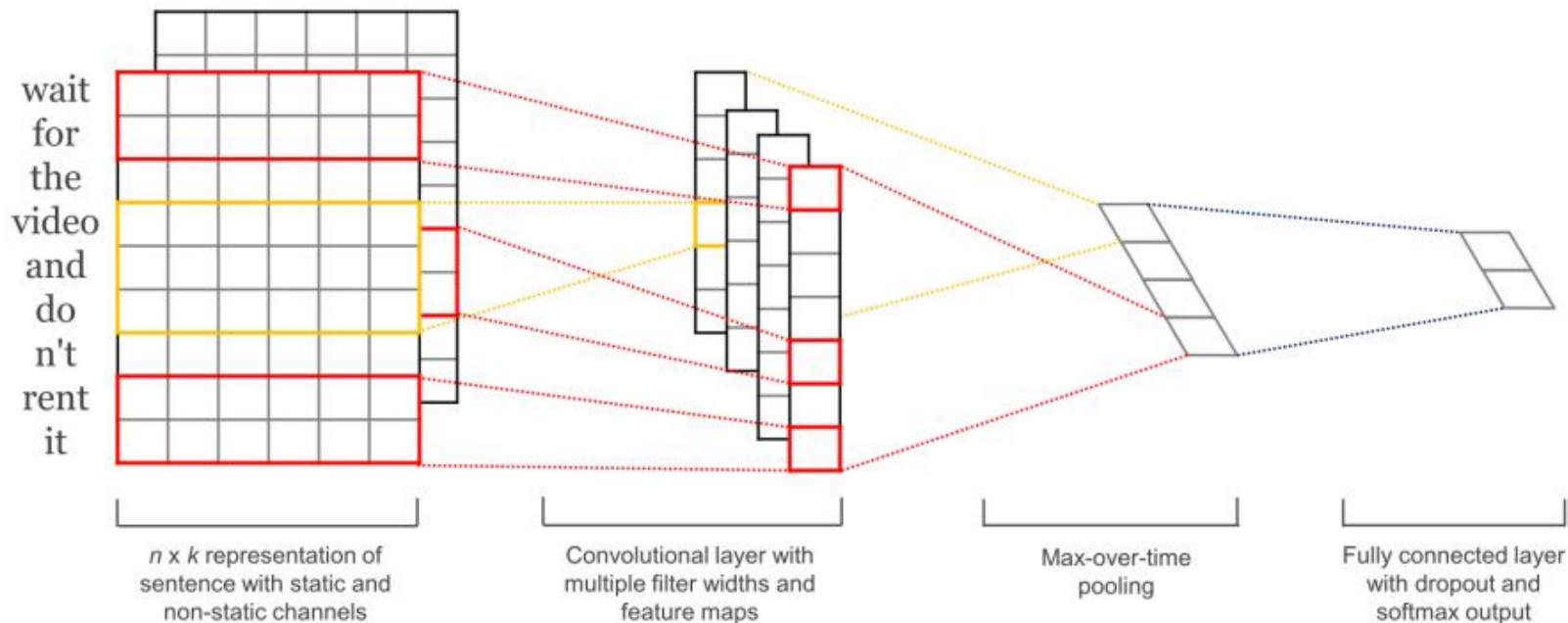
**Natural Language
Toolkit**



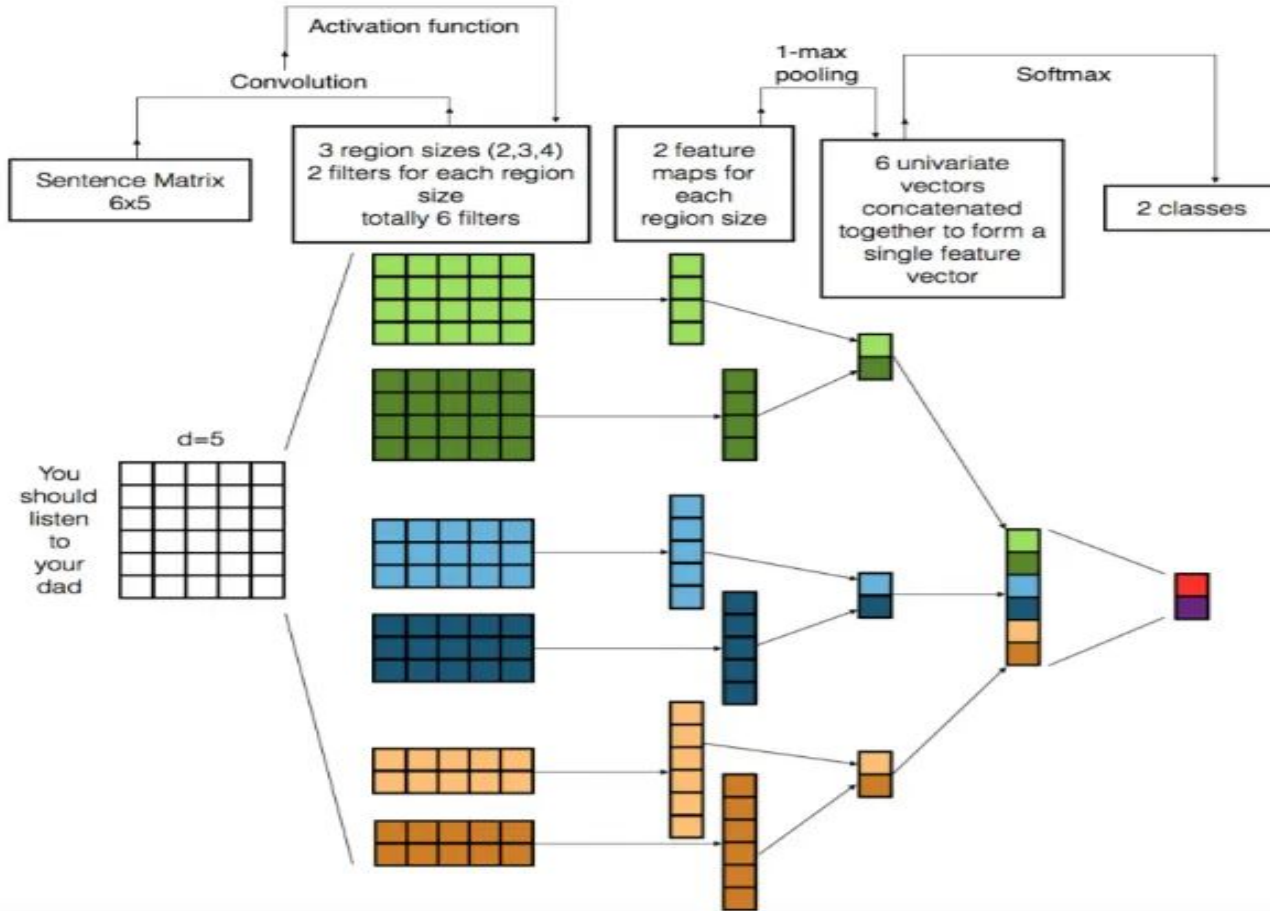
LANGUAGE MODELLING



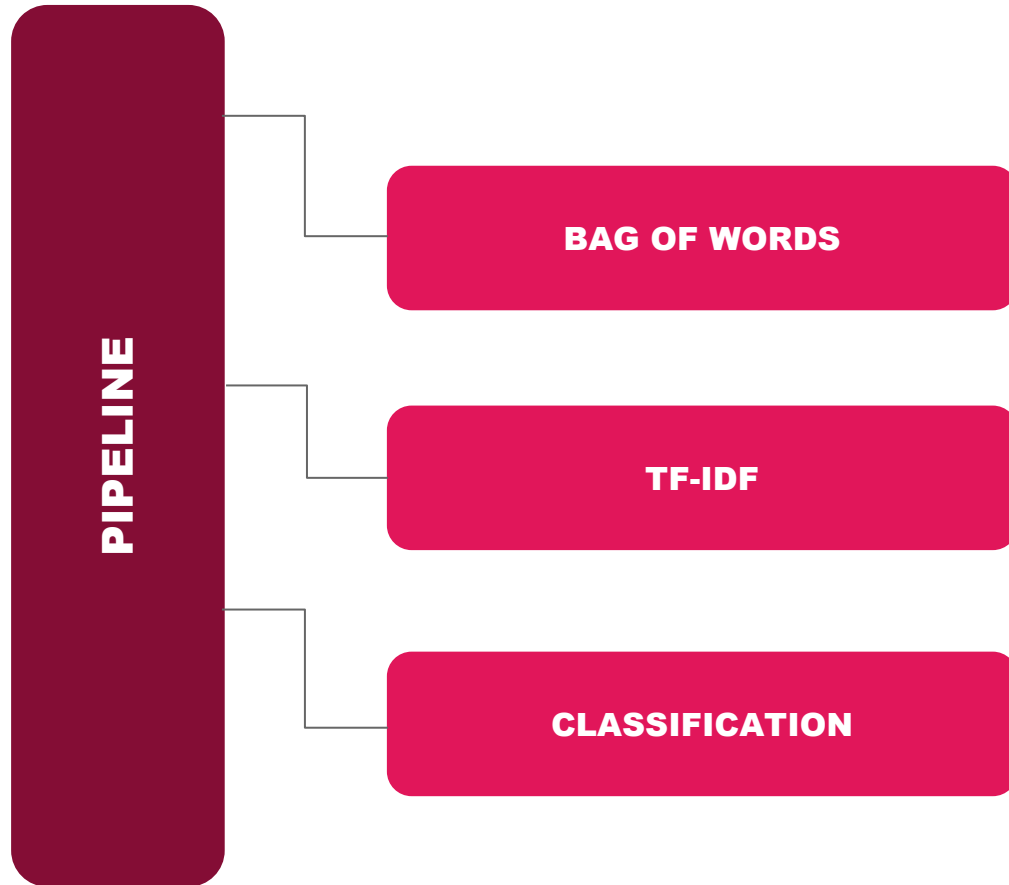
CNN FOR TEXT CLASSIFICATION



CNN - SENTIMENT ANALYSIS



PIPELINE



[illegible]

END OF CHAPTER