Natural language processing

## Natural Language

Natural language refers to the way we, humans, communicate with each other.

Natural Language Processing (NLP)

Natural language processing (NLP) refers to the branch of computer science and more specifically, the branch of [artificial intelligence](https://www.ibm.com/cloud/learn/what-is-artificial-intelligence) concerned with giving computers the ability to understand text and spoken words in much the same way human beings can.

Natural language processing Tasks

Several NLP tasks break down human text and voice data in ways that help the computer make sense of what it's ingesting. Some of these tasks include the following:

1. Speech recognition
2. Part of speech tagging
3. Word sense disambiguation
4. Named entity recognition
5. Co-reference resolution
6. Sentiment analysis
7. Natural language generation

### Text Preprocessing

1. Tokenization
2. Removing special characters
3. Convert sentence into lower case
4. Removing stop words
5. Stemming or Lemmatization

### Techniques to convert Text to Numerical Vectors

1. Bag of Words
2. TF IDF (Term Frequency - Inverse Document Frequency)
3. Word2Vec (by Google)
4. GloVe (Global Vectors by Stanford) - **Not Covered in this notebook**
5. Pretrained GloVe Embeddings
6. FastText (by Facebook) - **Not Covered in this notebook**
7. ELMo (Embeddings from Language Models) - **Not Covered in this notebook**
8. BERT (Bidirectional Encoder Representations from Transformer)

Vectorization / Embedding Techniques:

1. Bag of Word (BOW) ……..Counting
2. Term Frequency Inverse Document Frequency (TFIDF) ……..Counting
3. Word to Vector (W2v) ……..Deep learning
4. Global Vectors by Stanford (GLOVE) ……..Deep learning
5. Bidirectional Encoder Representation from Transformer (BERT) ……..Deep learning

 

 

 