## Natural Language Processing (NLP) — What is NLP ?

**Natural Language Processing (NLP) is** "*a branch of* ***artificial intelligence*** *that helps computers* ***understand***, ***interpret*** *and* ***manipulate*** *human language. NLP draws from many disciplines, including* ***computer science*** *and* ***computational linguistics***, *in its pursuit to fill the gap between human communication and computer understanding.*"

Also we can define NLP as "*a convergence of* ***artificial intelligence*** *and* ***computational linguistics*** *which handles interactions between machines and natural languages of humans in which computers are entailed to analyze, understand, alter or generate natural language.*"

And according to **Wikipedia**, Natural language processing (NLP) is *"a subfield of* ***linguistics***, ***computer science***, ***information engineering***, *and* ***artificial intelligence*** *concerned with the interactions between computers and human (natural) languages, in particular how to program computers to process and analyze large amounts of natural language data.*"

## Natural Language Processing (NLP) — Why NLP ?

NLP is about making computers **understand** how humans speak and **communicate** with each other in order to help the computers to engage in communication using natural **human language** in many forms, including but not limited to speech and writing.

We talk to each other in a number of different ways, whether it's through different grammatical structures or the regional idioms we use, so it's essential that NLP is in place so that this gap in communication can be bridged.

## Natural Language Processing (NLP) — Applications of NLP

Apart from personal assistants like Siri, Alexa and Google Assistant, some other applications of NLP include sentiment analysis, speech recognition and so on, so let's talk about some of it's applications :


The applications of Natural Language Processing (NLP)

### 1. Machine Translation

**Machine translation** (MT) refers to the **automatic translation** of a segment of text from one language to another. Humans may use MT

to help them render text and speech into another language, or the MT software may operate without human intervention.

MT tools are often used to translate vast amounts of information involving millions of words that could not possibly be translated the traditional way. The quality of MT output can vary considerably; MT systems require “training” in the desired domain and language pair to increase quality.

### 2. Speech Recognition

We are all familiar with **Google Assistant**, **Apple’s Siri**, **Amazon’s Alexa**, and **Microsoft’s Cortana**, and so on. all those virtual assistants AI technologies are built based on **Automatic Speech Recognition (ASR)** and **Natural Language Understanding (NLU)**, a procedure of converting speech into words, sounds, and ideas.

So **speech recognition** is simply the ability of a software to **recognize speech**. Anything that a person says, in a language of their choice, must be recognized by the software. Speech recognition technology can be used to perform an action based on the instructions defined by the human. The human needs to train the speech recognition system by storing speech patterns and vocabulary of their language into the system. By doing so, they can essentially train the system to understand them when they speak.

### 3. Sentiment Analysis

**Sentiment Analysis** (also known as **opinion mining** or **emotion AI**) is an interesting type of data mining that measures the inclination of people's opinions. The task of this analysis is to identify subjective information in the text.

Sentiment analysis helps to check whether customers are **satisfied** with goods or services. Classical polls have long faded into the background. Even those who want to support brands or political candidates are not always ready to spend time filling out questionnaires. However, people willingly share their opinions on social networks. The search for negative texts and the identification of the main complaints significantly helps to change concepts, improve products and advertising, as well as reduce the level of dissatisfaction. In turn, explicit positive reviews increase ratings and demand.

### 4. Text Classification

**Text classification** also known as *text tagging* or *text categorization* is the process of assigning a set of predefined categories to free-text or categorizing text into organized groups. By using Natural Language Processing (NLP), text classifiers can automatically analyze text and then assign a set of pre-defined tags or categories based on its content.

Text classification is becoming an increasingly important part of businesses as it allows to easily get insights from data and automate

business processes. There are many interesting applications for text classification such as **spam detection**, **email routing** and **sentiment analysis** etc...

### 5. Chatbots

**Chatbots** are applications that interact with users, usually over text, via websites or integrated into other platforms to ensure that customers get answers to their questions or find what they're looking for, in real time. Chatbots provide a conversational experience between the customer and the company without all of the extra steps of actually ringing up the customer service department.

The importance of NLP for a chatbot application is in making the customer feel understood in real time. Much like a human assistant, Natural Language Processing is what allows chatbots to understand and to dialogue with customers : it can ask questions and understand answers, identifying all the relevant information in the conversation in order to automate and effectively empower the customer assistance process.

### 6. Automatic Summarization

**Automatic Summarization** or **Text summarization** is the technique for generating a concise, meaningful and precise summary of voluminous texts (such as books, news articles, blog posts, research papers, emails, and tweets) while focusing on the sections that convey useful information, and without losing the overall meaning.

Automatic text summarization aims to transform lengthy documents into shortened versions, something which could be difficult and costly to undertake if done manually and its is one of the most challenging and interesting problems in the field of Natural Language Processing (NLP).

### 7. Spell Checking

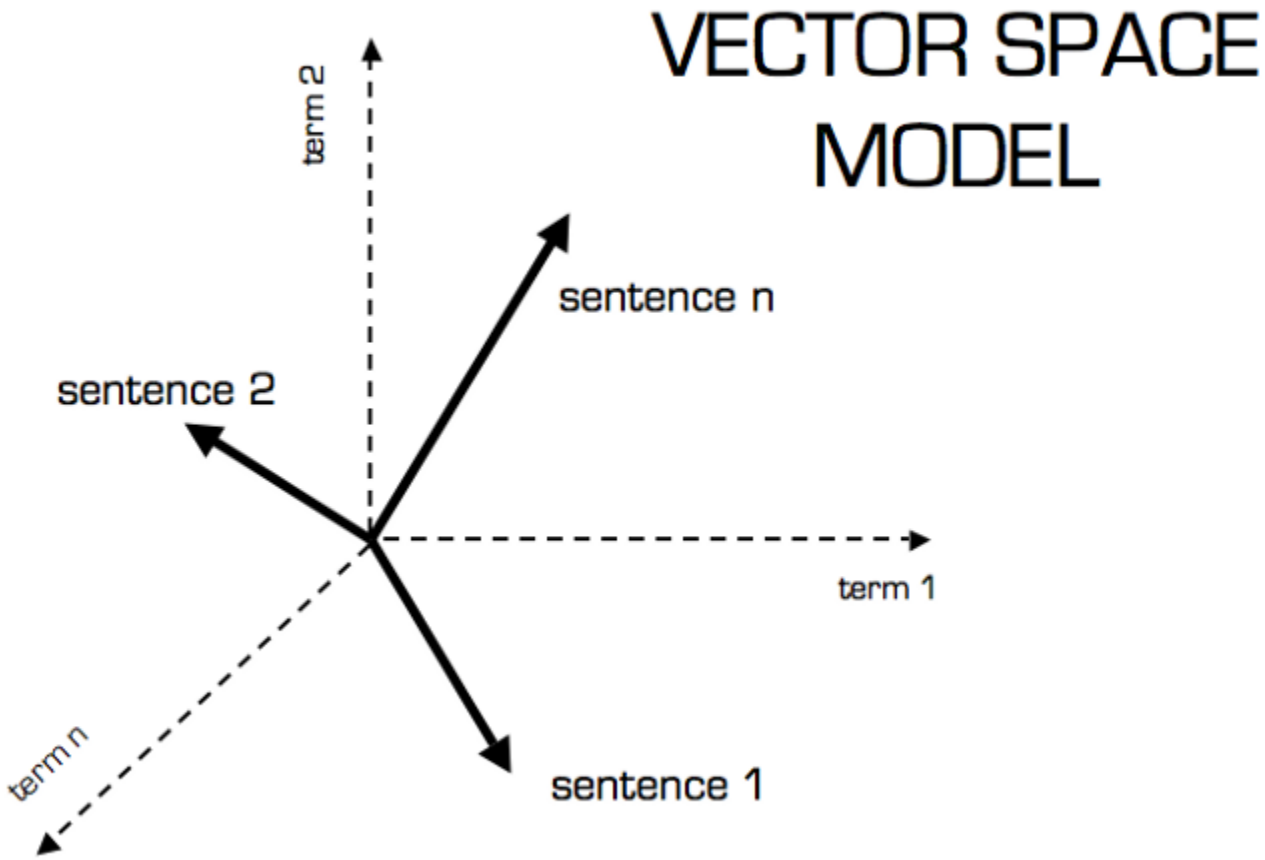
A **spell checker** or spell corrector is a technique which identifies the incorrect or misspelled words in a block of text and replaces them with the best possible combination of correct words. For find incorrect word firstly system checks the word in the dictionary. If the word is finding in the database then it assume to be correct word and if it is not present in the database then system assume this word incorrect and perform the required process to generate best possible combination of correct word.

Spell checking is one of the applications of Natural Language Processing that impacts billions of users daily, and most text editors (like Microsoft Word for example) or even search engines let users check if their text contains spelling mistakes.

What Is Natural Language Processing (NLP)?
Natural Language Processing (NLP) is a technique where text data like Mail, Social Media Posts, Web Page Content, SMS etc. is processed for extracting information from text data and use it in our computations and algorithm.

Text preprocessing is an essential step in building a Machine Learning model and depending on how well the data has been preprocessed, the results are seen.

In the vector space model, each word/term is an axis/dimension. The text/document is represented as a vector in the multi-dimensional space. The number of unique words means the number of dimensions.




When the dimension of data is very high i.e lot of unique words are there in corpus we might end up in curse of dimension and our document matrix will

be sparse. Have a look at this blog to know more about curse of dimension. So various text preprocessing steps are widely used for dimensionality reduction.