**Introduction to NLP**

Natural languages are the languages which have naturally evolved and used by human beings for communication purposes.

The term “natural” in the context of the language is used to distinguish human languages (such as Gujarati, English, Spanish and French) from computer languages i.e. formal language (such as C, C++, Java and Prolog).

**Definition:** Natural Language Processing (NLP) is a field of research and application that determines the way computers can be used to understand and manage natural language in the form of text or speech to do useful things.

NLP is the **scientific study of languages** from computational perspective.

NLP is a field of computer science and linguistics concerned with the interactions between computers and human (natural) languages.

NLP theoretically induced range of computational techniques for **analyzing and representing** naturally occurring text **at one or more levels of linguistic analysis** for the purpose of achieving human like language processing for a range of tasks or applications.

NLP has significant overlap with the field of computational linguistics, and is often considered a sub-field of artificial intelligence.

Natural language generation systems convert information from computer databases into readable human language.

Natural language understanding systems convert samples of human language into more formal representations such as parse trees or first order logic that are easier for computer programs to manipulate.

Many problems within NLP apply to both generating and understanding; for example, a computer must be able to model morphology (the structure of words) in order to understand an English sentence, and a model of morphology is also needed for producing a grammatically correct English sentence, i.e., natural language generator.

NLP is all about the processing of data.

for e.g. news articles, web pages, scientific articles, patents, emails, government documents, speech etc.

We have a lot of text data available but it is in unstructured format.

Data is not available in single language for example English

Popular languages in internet are **English, Chinese, Spanish, Arabic, French, Others**

We will be not able to read all the languages. Therefore we need a tool than can take any language and convert into a language that we can understand.

Even instead of converting all text, we are interested in summarization and then conversion.

**Fundamental and scientific goal: Deep understanding of broad language.**

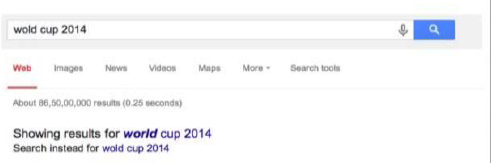
**Can we teach computers to understand natural language and then human machine interaction is possible in natural language.**

**Engineering Goal: Design, implement and test systems that process natural languages for practical applications.**

**NLP Goals can be very ambitious and practical**

**Some of the Applications of NLP**

**1) Auto Completion**

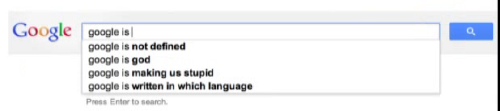
****

Suppose in Google, you type a query "wold cup 2014", and misses our r in world. Then Google will give you some sort of reply that Ok, are you looking for world cup 2014.

So, instead of wold did you mean world?

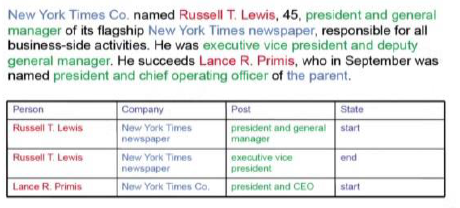
In NLP, Automatic query correction problem can be solved in very systematic manner

**2) Query prediction and Completion in search engines**



If you type a start typing in a query, then search engine will try to predict what is the complete query that that you are planning.

**3) Information Extraction**

****

We have a lot of unstructured data in the sense of news report and other materials.

We want to identify what are the entities of interest and what are the various relations between these entities.

For example, We can identify that Russell is a person who works on the post of president and general manager in the company New York times newspaper and he just started his post at this movement. So, this information is available in the text data, but in a unstructured format.

Information extraction convert unstructured data to structured format.

**4) Domain Specific Chatbots**

Response from machine TA is observed by human TA. Human TA made correction if any. Initially responses from machine TA was not good. But after some training, machine TA was too good to handle domain specific problems.

You want to buy some product on an e-commerce website, instead of having to search everything. We can just provide your specifications to the chatbot and it can search a product for you.

In the case of any flight booking system or banking system where you can give your queries and the chatbot can come up with the possible reply by looking into the document and so on. And this is the practical application that can be solved using NLP.

**5) Sentiment Analysis**

From tweets, opinions and comments that we provide in social media, we can find out what are various sentiments of users and with that we can also find out are they some transitions and sentiments of the users over the years.

**6) Spam Detection**

In Gmail or any other the web service, many of the emails going to the spam folder directly without even bothering to us. you. Once and email comes in, so the system tries to see is it spam or not by doing again text analysis over there and if it just spam it is not even shown to you in your inbox it is directly send to some a spam folder. Spam deduction is again a very practical goal in not only in your emails also on with social media even on tweets. YouTube comments, even YouTube videos finding out what are its spams is again very interesting and challenging problem.

**7) Machine Translation**

Opening a web page from some other country or suppose we are going to visit countries like China or Japan and we want to read web page which is their language. We can use Google translate to load that page in English or in any other language for you.

**8) Text summarization**

Given a big news article or scientific article, we can summarize that in short.