

# Natural Language Processing

Course Code: CSET346

---

Course Instructor:  
**Dr. Ghazaala Yasmin**

# Question????

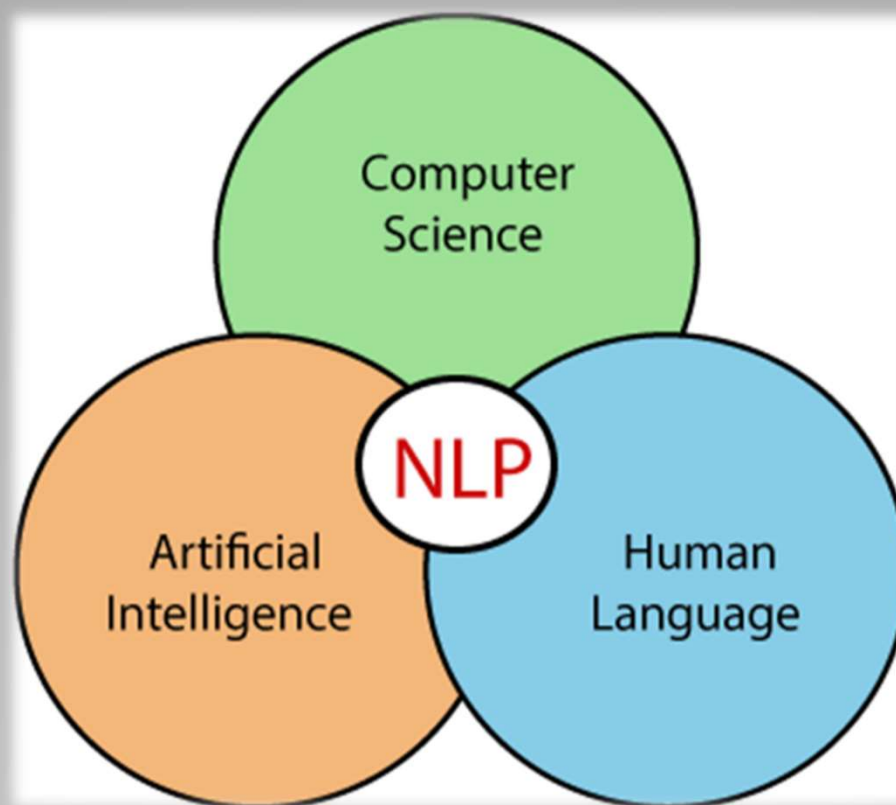
Have you ever used any NLP products/ NLP Powered tools ?



# What is NLP?

- A sub-field of Artificial Intelligence (AI)
- An inter disciplinary subject
- Aim:
  - To build intelligent computers that can interact with human being like a human being !!

# At Glance...



- ❑ Understand language analysis & generation
- ❑ Communication
- ❑ Language is a window to the mind
- ❑ Data is in linguistic form





# Course Syllabus:

Sno.	Lecture	Sno.	Lecture
L1	Introduction	L18	Neural Machine Translation
L2	Representation of word, Sentence	L19	Seq2Seq Modelling
L3	Word embedding	L20	Attention
L4	Word Senses	L21	Project discussion
L5	Linguistic Structure: Dependency Parsing	L22	Question Answering Bot
L6	Word Window Classification	L23	1D-CNN for NLP
L7	Neural Networks for text	L24	Sub-word Models
L8	N-gram Language Models	L25	Contextual Representations
L9	Recurrent Neural network	L26	Transformers
L10	Vanishing Gradients	L27	Self-Attention For Generative Models
L11	LSTM (Long sort term memory)	L28	Natural Language Generation
L12	GRU (Gated recurrent Unit)		
L13	Part of speech tagging		
L14	BERT		
L16	XLnet		
L17	Statically Machine Translation		

# Course Evaluation

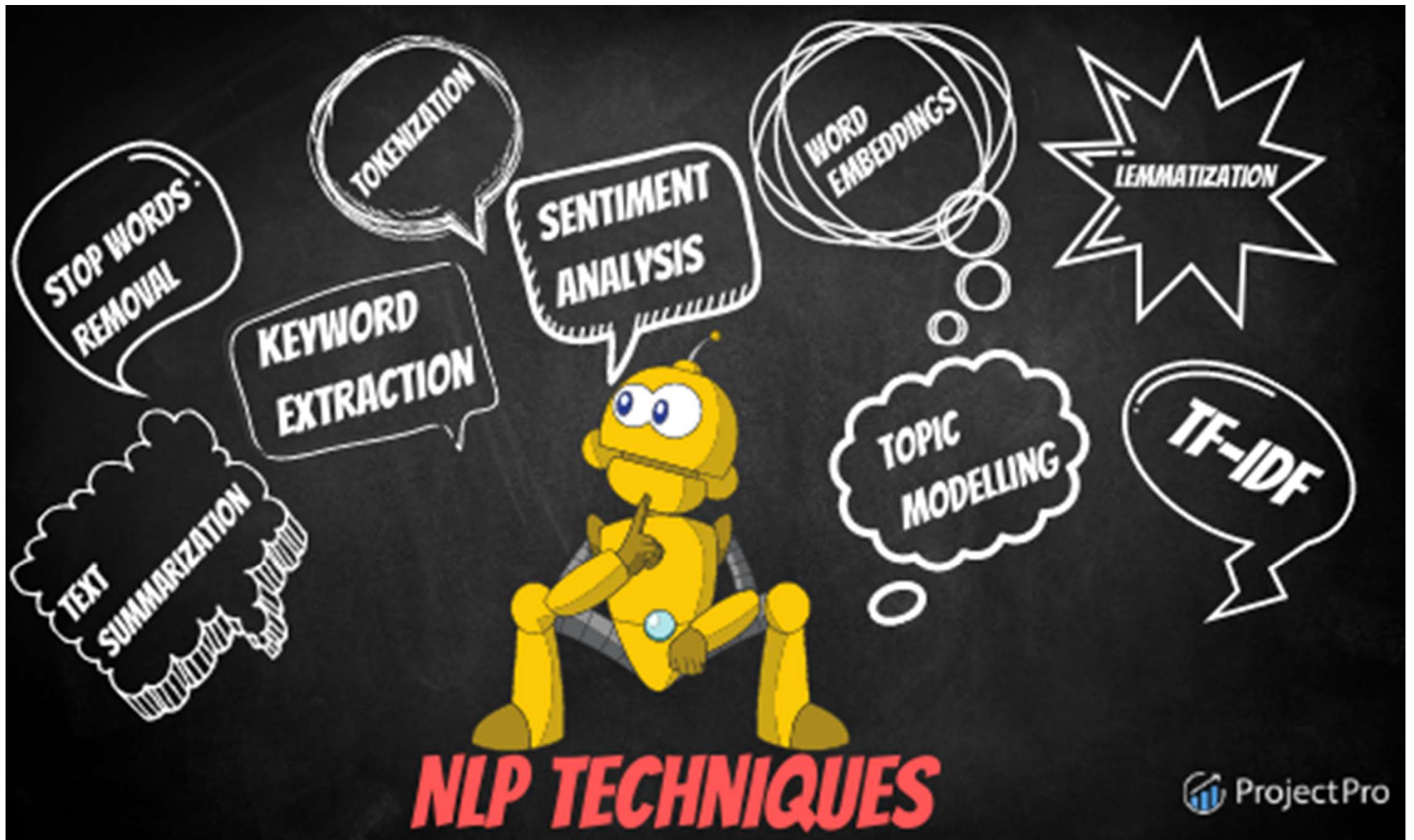
Components	Distribution
End Term Semester Examination	35
Continuous evaluation	Quiz (5)
	Certification (30)
	Lab Assessment (30)

# Course Certification

<https://www.coursera.org/specializations/natural-language-processing>

- ❖ Natural Language Processing with Classification and Vector Spaces
- ❖ Natural Language Processing with Probabilistic Models
- ❖ Natural Language Processing with Sequence Models
- ❖ Natural Language Processing with Attention Models

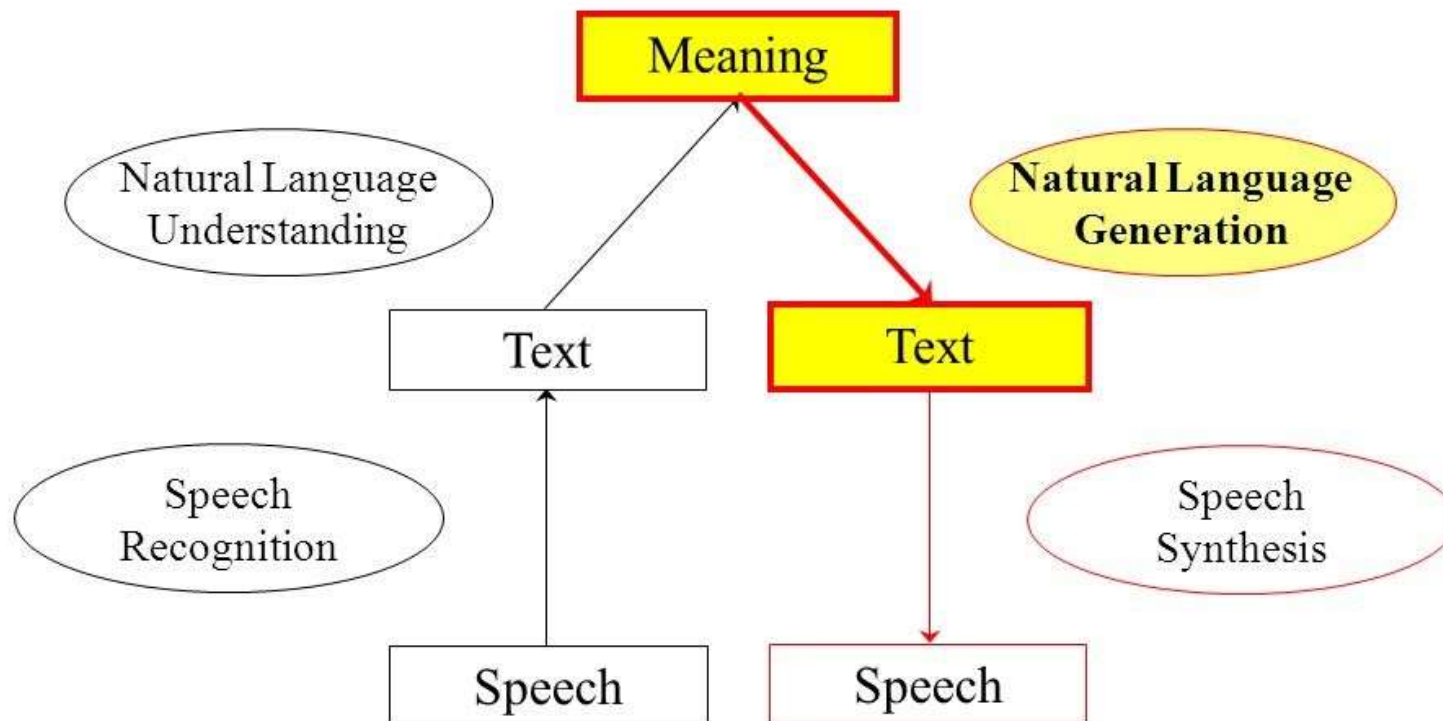




# Prepare your laptop for NLP

- Install following:
  - Python
  - Jupiter
  - TensorFlow and Pytorch
  - NLTK
  - Gensim
  - Polyglot
  - TextBlob
  - Download Glove
  - Download Word2vec

# Language Technology



# Corpus

- **A collection of authentic text or audio organized into datasets.**
- **Authentic means text written or audio spoken by a native of the language or dialect.**
- **A corpus can be made up of everything from newspapers, novels, recipes, radio broadcasts to television shows, movies, and tweets.**



# Application of NLP



## Application of NLP

### More Deeper Application of NLP

#### Group 1

Cleanup, Tokenization

Stemming

Lemmatization

Part of Speech Tagging

Query Expansion

Parsing

Topic Segmentation and  
Recognition

Morphological Degmentation  
(Word/Sentences)

#### Group 2

Information Retrieval and  
Extraction (IR)

Relationship Extraction

Named Entity Recognition  
(NER)

Sentiment Analysis/Sentence  
Boundary Disambiguation

World sense and  
Disambiguation

Text Similarity

Coreference Resolution

Discourse Analysis

#### Group 3

Machine Translation

Automatic Summarization/  
Paraphrasing

Natural Language Generation

Reasoning over  
Knowledge Based

Quation Answering System

Dialog System

Image Captioning & other  
Multimodel Tasks

# Corpus

- A corpus may contain texts in a single language (*monolingual corpus*) or text data in multiple languages (*multilingual corpus*).
- Multilingual corpora that have been specially formatted for side-by-side comparison are called *aligned parallel corpora*.

# Task

- Prepare corpus for your mother language
  - Monolingual
  - Multilingual