# Natural Language

- English, Hindi, Bengali...
- C++, Java?--- Formal language

# NLP?

Natural language processing (NLP) is the ability of a computer program to understand human language as it is spoken or written

Moumita Basu, Department of CSE, UEM Kolkata

# Motivation of Learning NLP?

- Huge amount of data in internet – in form of texts (Natural Language)
- Documents, Emails, Blogs, Social Media post
- Appications needed to process this huge amount of unstructured data

# Relation between NLP and Text mining

- Natural language processing is responsible for understanding meaning and structure of given text.
- Text Mining or Text Analytics is a process of extracting hidden information inside text data

# Applications

- Classification of text
- Sentiment analysis
- Automatic Question Answering
- Automatic  Summarizations
- Recommendation systems
- Information Extraction (Structured information from email like calendar)
- Machine translation.

Moumita Basu, Department of CSE, UEM Kolkata

# Applications

- A software program capture the sound of lecture delivery  in a class and convert it to  textual file which can eventually uploaded
- Software programs which will navigate the web, and give to the organization peoples feedback
- E-governance ,  Disaster Management

# Informal variation of words (Noisy)

| Variations | Excerpts from tweets |
|---|---|
| **Variations in spellings of a word** | |
| epicentre<br>epicenter | 6.7 magnitude #earthquake epicentre 49 km from Banepa in #Nepal says USGS<br>5.0 earthquake, 29km SSW of Kodari, Nepal. Apr 26 13:11 at epicenter |
| sindhupalchowk<br>sindhupalchok<br>sindupalchowk | #Sindhupalchowk 1100+deaths and 99% Houses are Down<br>Indian national Azhar 23, missing. Last location Sindhupalchok. Plz help.<br>Food Distribution in sindupalchowk, sufficient for 7 days for 500 victims |
| **Arbitrarily shortened forms of a word** | |
| building<br>bldg | Earthquake destroyed hospital, road, building in Kavre district of Nepal.<br>Nepal quake stresses importance of earthquake resistant bldg designs in entire NCR. |
| medical<br>med | India: NDRF personnel, crack medical team with relief rushed to #Nepal<br>4 planes to leave for #Nepal tmrw carry meds, med team, 30-Bed Hospital |

# Ambiguous !!

- Need-tweets
  - Chirayu hospital has been evacuated after cracks after #NepalQuake , ishan hospital treating patients on the road
  - Emergency Alert Please support antibiotics and IV sets for earthquake victims being treated in Gandaki Medical College
  - Water Scarcity at Valley !!! every hands wants at least one bottle drinking water !!! #earthquake #NepalQuake [url]
  - There is shortage of Blood as well as oxygen cylinders...Nepal is in huge crisis
- Availability-tweets
  - Dr. Madhur Basnet leading medical team going to remote villages of Gorkha dist which was epicenter of earthquake. His cell: [number]
  - Govt launches national disaster helpline 1078 #Earthquake #Nepal #India [url]
  - Round the clock - our medical team treats earthquake victims #IsraelInNepal [url]

# AI and NLP

- Machine is considered to be intelligent if it understand, interpret and speak natural language
- The objective of NLP is to make computer/machines as intelligent as human beings in understanding language
- Natural language comes in two different forms:
  - one is the spoken utterances
  - other one is written communication

# AI and NLP

- AI: Demanding machine should come closer to human being
- NLP: Computer being able to process human languages like English, Hindi
- NLP is forcing function for AI

Moumita Basu, Department of CSE, UEM Kolkata

# Computer Lack knowledge

- No common sense knowledge
- No reasoning capacity
- No experience

Unless we teach

Moumita Basu, Department of CSE, UEM Kolkata

# AI perspective of NLP (pre-web)
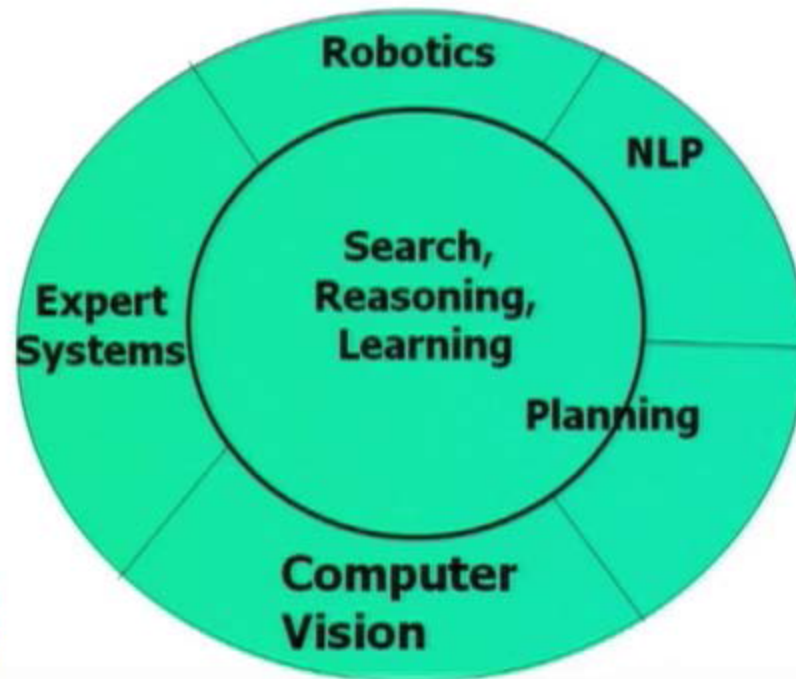
# Layers of AI



Perpectivising NLP: Areas of AI and their inter-dependencies

Moumita Basu, Department of CSE, UEM Kolkata
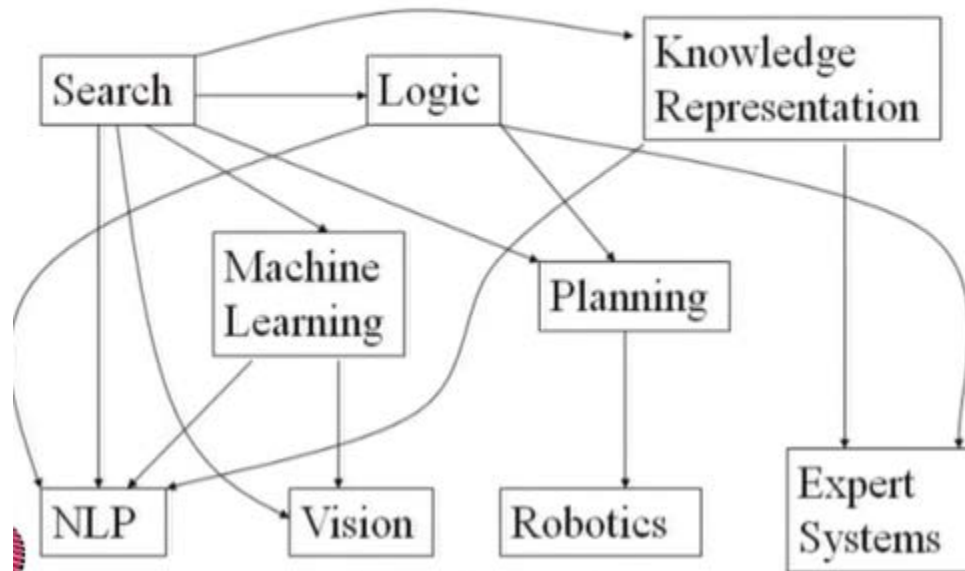
# Layers Of AI

- First layer: Search, logic and knowledge representation
- Second layer: Machine learning and planning
- Final layer: Natural language processing, vision, robotics and expert systems

# Planning

Planning is arranging a sequence of actions to achieve a goal

Sentence is containing knowledge
knowledge has to be extracted and embedded in the machine

# Search

- "I went to the bank to withdraw some money" ---- Ambiguity
- A program will have to read the sentence left to right

Moumita Basu, Department of CSE, UEM Kolkata

# Logic

- Textual knowledge has to be converted into logical forms which a machine can process
- proposition calculus, predicate calculus and some forms of non monotonic logic are used for natural language processing

Moumita Basu, Department of CSE, UEM Kolkata

# Machine Learning

- The internet is full of textual documents

 How much of data can human being possibly see?

# Machine Learning

- Machine learning techniques, statistical techniques which look at the data and obtain the knowledge content of the data
- Machine learning is feeding into natural language processing

# Other perspective of AI

- Robotics:  Embedded s/w inside the robot enabling it to perform human like actions

- Vision:  Ability of a computer to process a scene

- Expert System: Expert level performance , like emulating behaviour of experts like doctors

# Goals and applications of NLP

- Science Goal:  Understand how we process and generate language
- Engineering Goal:  Use NLP to solve real life problem ( creates tools & resources)

# AI perspective of NLP(post-web)

- Strong presence of Information Retrieval
- Machine learning contributed a lot to Information Retrieval

# AI perspective of NLP(post-web)



Moumita Basu, Department of CSE, UEM Kolkata
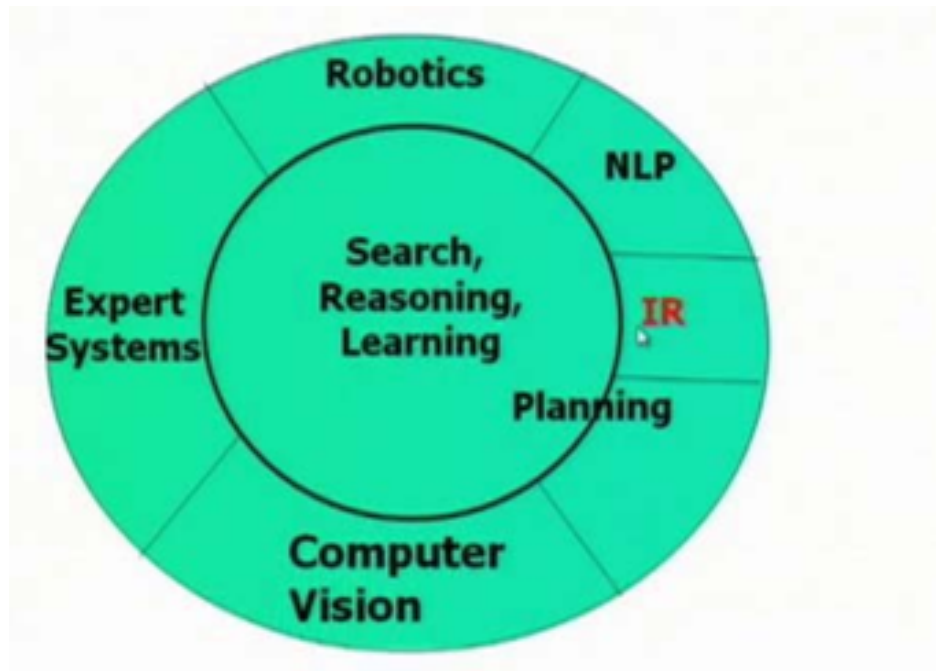
# IR

User satisfaction with respect to his or her information need (Query) Depends on

- Ranking: Ranked according relevance
- Coverage: How much of the information we have been able to capture

Moumita Basu, Department of CSE, UEM Kolkata

# DIFFERENCE BETWEEN NLP, AI, ML, DL & NN

**AI or Artificial Intelligence** – Building systems that can do intelligent things.

**NLP or Natural Language Processing** – Building systems that can understand language. It is a subset of Artificial Intelligence.

**ML or Machine Learning** – Building systems that can learn from experience.

**NN or Neural Network** – Biologically inspired network of Artificial Neurons.

**DL or Deep Learning** – Building systems that use Deep Neural Network on a large set of data. It is a subset of Machine Learning.