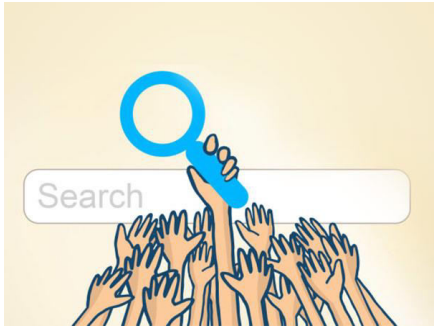
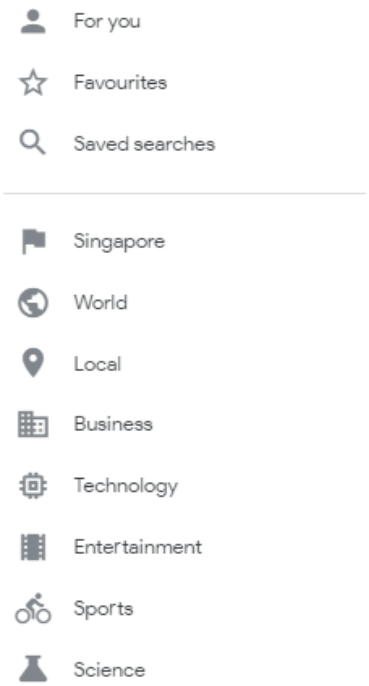
Natural Language Processing (NLP)

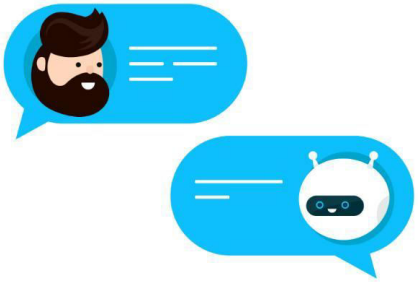
NLP Applications



Search Engines



News Categorization



Conversation Systems

About the Course

- Introduction to text data and natural language processing
- Key techniques related to natural language processing
- Implement natural language processing on real datasets
- Approaching text classification problems
- Deep Learning for natural language processing
- 3 Hands-on projects

Natural Language Processing : Introduction

Module Outline

1. Dataset Types : Structured vs Unstructured
2. Unstructured dataset : Text data
3. What is natural language processing
4. Examples and business use cases

Dataset types : Structured and Unstructured

Structured Dataset

- A dataset having fixed number of dimensions
- Tabular data or Key value pairs

Emp_id	Emp_name	Designation	Salary	City
1001	RAHUL	Technical Leader	63k	LUCKNOW
1002	KARAN	Junior Developer	42k	JAIPUR
1003	SEEMA	Junior Developer	45k	AGRA
1004	SHREYA	Senior Developer	50k	KANPUR
1005	REET	Project Manager	86k	AMRITSAR
1006	GAURAV	Technical Leader	67k	DELHI

Place	Maximum Temp in °C	Minimum Temp in °C
Pitampura	32.5	25.4
Akshardham	31.1	25
Yamuna Sports Complex	32.8	25.3
Delhi University	32.2	24.7
Safdarjung	29.7	23.3
Palam	31.9	24.9
Ridge	31.2	21
Lodhi Road	29.4	24
Noida	32.8	25

```
{
  "id": 123,
  "name": "Sandeep",
  "subject": "Computer",
  "marks": 23
},
{
  "id": 193,
  "name": "Raja",
  "subject": "Mathematics",
  "marks": 25
},
{
  "id": 223,
  "name": "Smith",
  "subject": "Geography",
  "marks": 20
},
```

Dataset types : Structured and Unstructured

Unstructured Dataset

- A dataset having no fixed dimensions
- Forms : audios, videos, images, text

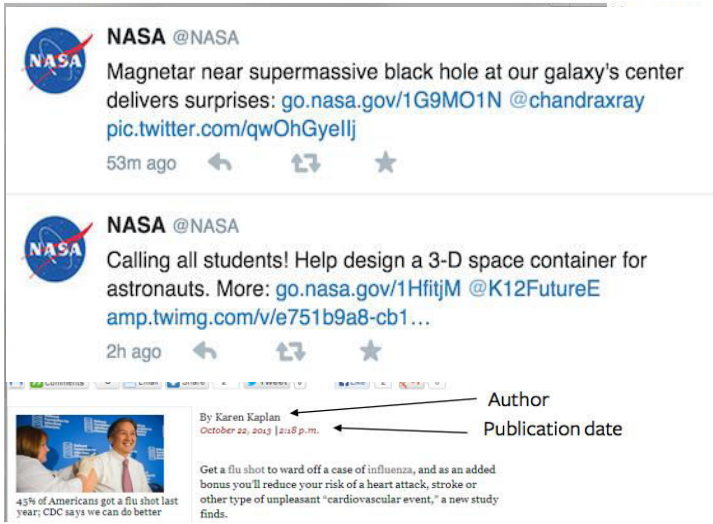
Cats

Dogs

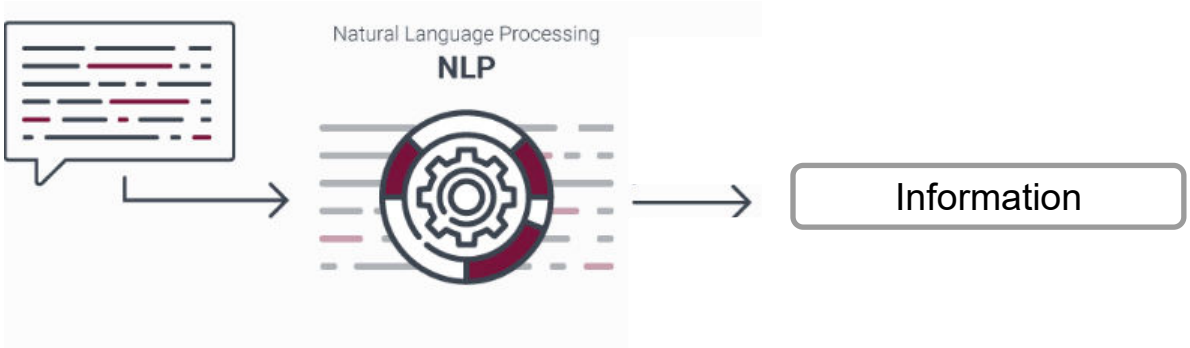
□□szzÙ□˜IDATx□Ìî□ê$Y□EO¢≤‹Ó±ÌôòY€ð□X€ªÅU`m+]]ˆøøΩÊÿð››YJÁ{˚ƒbm
≥Nƒâ˚ÖÚfi"5j‘®Q„□èΔ◊□–˜qª□™´;fô‘Üƒ±ÏSÉú□xf˜LÜˆ0
-)lKcn˜TÍ˜πð˝4` Á□7ç□~Áô[]ekÀÑÜ˙«„T˙†óã%BvB7□ËöA□□ú—/»□
=–ÄP□4;%ˆÒKÖ%ð)6□¸w□pÂ–!|öYû7r□À˙ÔJT™Á□©·T*xÆK□ÜH)ë¿ç˝˙rgŸ%
£fiâ\æû\]=V–,êuÀòπqvΩ+=˜R@Ú-0gΣ∑sÎaámìÓŸs□≤Ÿë3÷≠;i"§âô˜ï*àä“˜âÇ□□
ÑÆ□: ç*□ùXJ¢(fÛñ5t◊uZLÛ,ødåÒ£z˜>X 7†M□5,"‹...
ß‹j&□ËÜÅ©tmõÖ[∑□frlìœ £□$UÑ□8j‘€k◊”£•ÖnæáàB‹ P□°{□-
≠≠§ª˜¿ÎËÿ□:Œº¿qÊ□~“LØXú
Ã˜ÏÄÜ¸√fiNªúì0;ÀML+A*i$S)À□□#u□]◊ëq^˝□˜™(•□K!□™XΔ1(ÂáJj†&□Q□U?

Text dataset

- Written form of any language
- Characters or words arranged together in a meaningful and formal manner
- Grammar rules and defined structures
- Examples:
  - I. Social Media : tweets, posts, comments
  - II. Conversations : messages, emails, chats
  - III. Articles : news, blogs, transcripts



Natural Language Processing (NLP)



- Branch of data science that deals with deriving useful information from the text data
- Analyzing, understanding, and utilizing text data
- Applied NLP : designing and building applications that enable interactions between machines and natural languages

Business Use Case 1

Finding audience insights about smartphone

A mobile brand wants to increase the sales of their smartphone model

NLP Solution :

- Monitor social media data
- Identify most talked issues
- Understand audience sentiments
- Improve issues and causes of negative sentiment

Business Use Case 2

Automatic categorization of customer queries

A telecom company wants to reduce the time taken to resolve customer queries

NLP Solution :

- Monitor the query text – top keywords and phrases
- Map the keywords to the descriptions of respective departments
- Model for automatic categorization of queries

Business Use Case 3

Identify the patients at risk of cancer

A hospital wants to create a model to early detect the patients at risk

NLP Solution :

- Analyze the history of patients and drugs prescribed.
- Identify the entities from their history
- Business rules to classify the risk levels of patients