GROUP NO. 01

GROUP MEMBERS:

Sanjana Asrani 01

Abhishek Chhabria 15

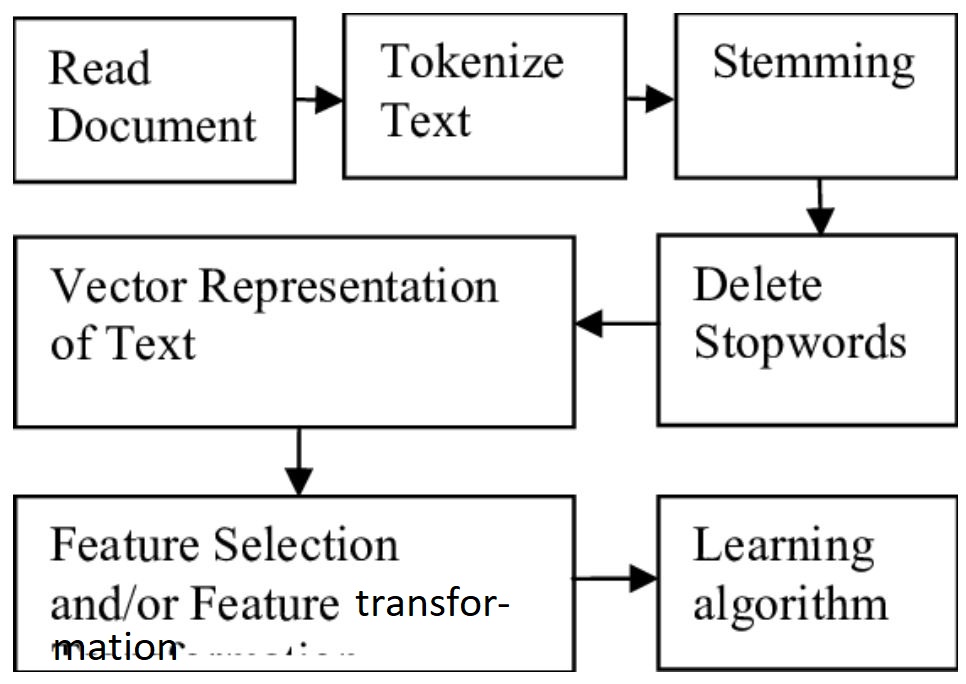
Jayesh Dayalani 20

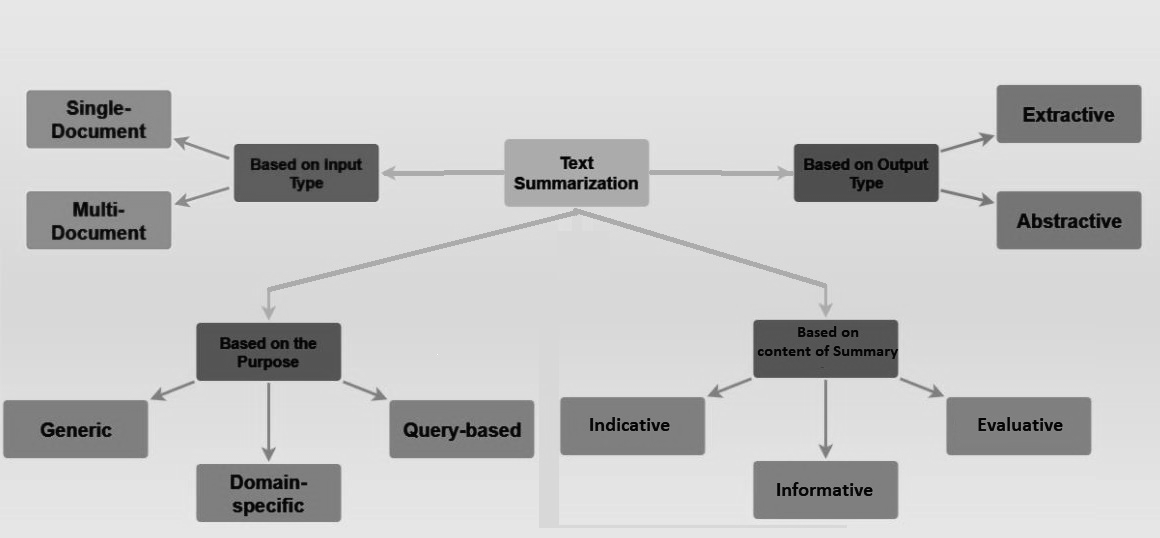
**Aim**: Study various applications of NLP and Formulate the Problem statement for Miniproject based on a chosen real world NLP application.

**Theory**:

Natural Language Processing (NLP) is a field of artificial intelligence that focuses on the interaction between computers and human language. NLP has a wide range of applications across various domains, revolutionizing the way we interact with and analyze text data. Here's a brief overview of some key applications of NLP:

1. **Text Classification:** NLP is used for classifying and categorizing text documents into predefined categories. It's widely used in spam email detection, sentiment analysis for social media, and content recommendation systems.





2. **Sentiment Analysis:** NLP helps determine the sentiment (positive, negative, or neutral) expressed in textual data. This is valuable for understanding customer opinions, brand reputation, and social media monitoring.

3. **Language Translation**: NLP powers machine translation systems like Google Translate, allowing users to translate text from one language to another. This has numerous applications in global communication and content localization.

4. **Chatbots and Virtual Assistants:** NLP-driven chatbots and virtual assistants are used for customer support, automating responses, and providing information and services through natural language conversations.

5. **Information Retrieval**: Search engines like Google use NLP to understand user queries and retrieve relevant search results. NLP also plays a role in document summarization and content extraction. Information Retrieval (IR) is finding material (usually documents) of an unstructured nature (usually text)

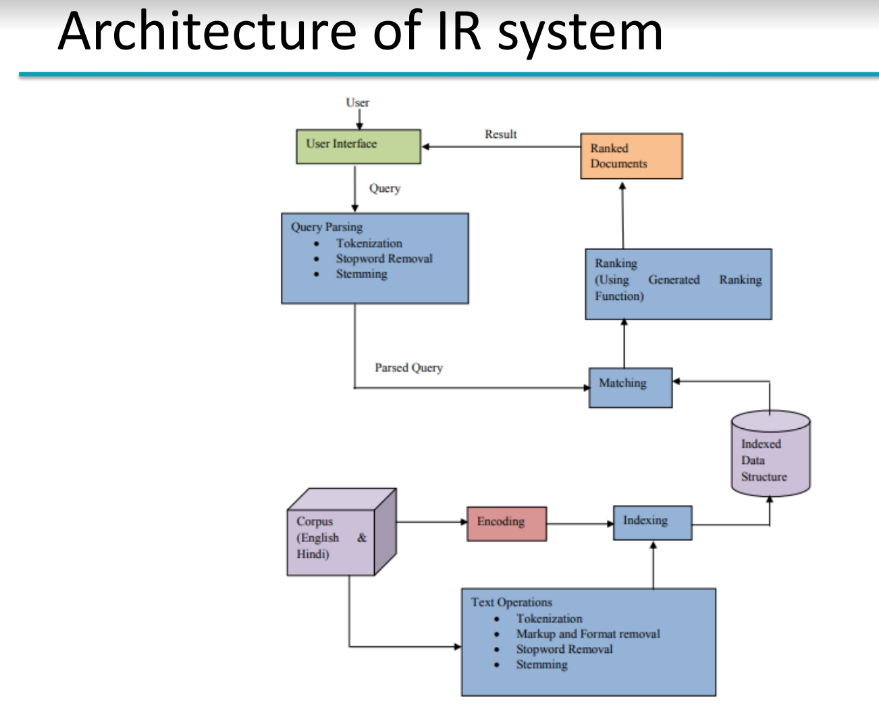
that satisfies an information need from within large collections (usually stored on computers).

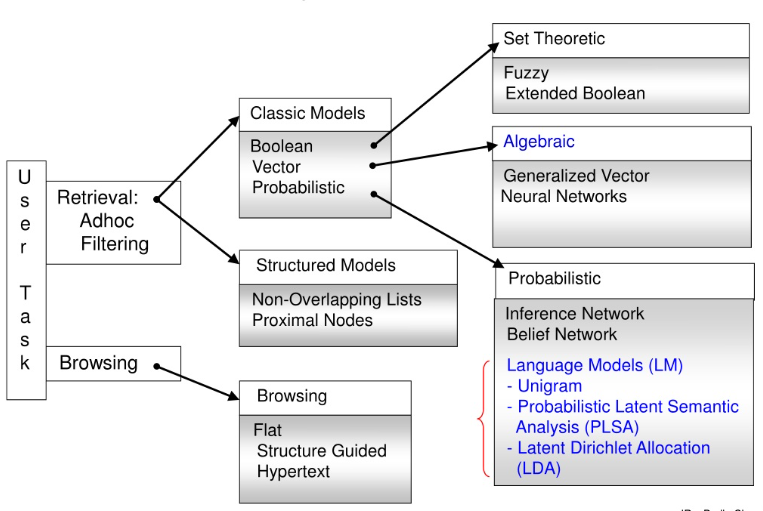
▪ These days we frequently think first of web search, but

there are many other cases:

▪ E-mail search, Searching your laptop

▪ Corporate knowledge bases, Legal information retrieval





6. **Speech Recognition**: NLP technology enables speech recognition systems like Siri and Alexa to convert spoken language into text, facilitating voice commands and dictation.

7. **Text Generation**: NLP models like GPT-3 can generate human-like text, making them valuable for content generation, writing assistance, and creative applications like poetry and storytelling.

8. **Named Entity Recognition (NER)**: NLP can identify and classify named entities such as names of people, organizations, locations, and dates within text. This is crucial for information extraction and knowledge graph construction.

**Problem Statement:**

To develop an NLP-based model that summarizes the news articles and translates it to regional language.

**Approach**:

1. Preprocess the news articles using libraries like spaCy or **NLTK** to clean and tokenize the text.
2. Use **TF-IDF** to identify important keywords or phrases in the articles.
3. Employ cosine similarity to measure the **similarity** between sentences or paragraphs, helping to identify relevant content for summarization.
4. Implement a summarization model (e.g., BERTSUM or TextRank) to generate summaries based on the identified content.
5. Utilize a machine translation model (e.g. Transformer-based) to translate the summary from the source language to the regional language.

**Conclusion:**

Thus Real life applications of NLP have been studied about with some of them being already even implemented during labs for eg. NER (Named Entity Recognition). The topic for the mini project has been chosen and formulated accordingly.