

**Natural Language Processing**

***(Mini Project Synopsis)***

**Group Members:**

***TY BTech CSE – Panel C – Batch C1***

1. Aditi Singh – 1032221947 – 09
2. Gaganjot Kaur Badwal – 1032222170 – 29
3. Nimisha Manna James – 1032222171 – 30



**Title**: Document clustering and summarization website

**Abstract:**

In a time of rapidly increasing digital content, managing and comprehending significant amounts of unstructured text data presents a major challenge. This initiative suggests an online platform for clustering and summarizing diverse documents, utilizing Natural Language Processing (NLP) methods to automatically categorize and condense documents according to content similarity. The platform enables users to upload several documents in DOCS format which are subsequently analyzed using text extraction, vectorization, and clustering methods. The documents are organized into separate folders according to their semantic similarities, making sure that related content is classified together.

After clustering is finished, the system creates an extra summary file for every cluster, emphasizing the shared themes, important similarities, and distinctive differences among the documents in that category. This allows users to swiftly understand the connections between documents in a cluster and what sets each apart. The suggested approach utilizes machine learning, deep learning, and NLP models for analyzing text, embedding documents, and clustering, guaranteeing high precision in similarity identification. The system offers an effective method for researchers, analysts, and professionals to structure, examine, and derive significant insights from extensive groups of diverse textual information.

**Objectives:**

* Automated Document Clustering – Categorize heterogeneous documents into meaningful clusters based on content similarity.
* Summarization with Similarity & Differences Analysis – Generate summaries highlighting common themes within each cluster and unique aspects of individual documents.
* User-Friendly Web Interface – Provide an intuitive platform for document upload, clustering, and retrieval with easy access to summaries.

**Literature Survey:**



**System Flow Diagram:**

