Overview:

In the dynamic realm of communication, email correspondence has emerged as a pervasive medium for both professional and personal interactions. With the sheer volume of daily email exchanges, there is a growing demand for efficiency, consistency, and standardization in email content. This research explores the application of Natural Language Processing (NLP) techniques to automate email standardization and facilitate template generation. The primary aim is to boost communication efficiency, alleviate cognitive load, and elevate the overall quality of interactions within email-based communication systems.

Literature Review:

While email communication plays a crucial role in modern business interactions, the manual effort invested in composing and responding to emails poses a significant productivity bottleneck. This research addresses this challenge by harnessing NLP techniques to automate email standardization. By establishing a uniform format and structure, this approach seeks to streamline communication processes and enhance clarity in conveying information. Despite previous research exploring various facets of email communication, the specific focus on standardization and template generation within the realm of NLP applications remains relatively unexplored. This research builds upon existing work in email analysis, drawing inspiration from studies in text mining, machine learning, and natural language understanding.

Methodology:

The proposed methodology involves a comprehensive multi-step process. Initially, a diverse dataset of emails is collected and preprocessed to ensure relevance and quality. Text mining techniques are then applied to extract key features such as common phrases, greetings, and sign-offs. NLP models, including Named Entity Recognition (NER) and sentiment analysis, are employed to comprehend the context and tone of emails. Based on these insights, machine learning algorithms are utilized to generate templates, enabling the automatic creation of standardized responses.

Key Components:

Text Mining and Feature Extraction:

Utilizing text mining techniques to identify recurring patterns and key elements within the email content.

Extracting common phrases, salutations, and other linguistic features to inform the template generation process.

NLP Analysis:
Applying NLP models for sentiment analysis to discern the emotional tone of emails.

Implementing Named Entity Recognition (NER) to identify entities, such as names, dates, and locations, crucial for contextual understanding.

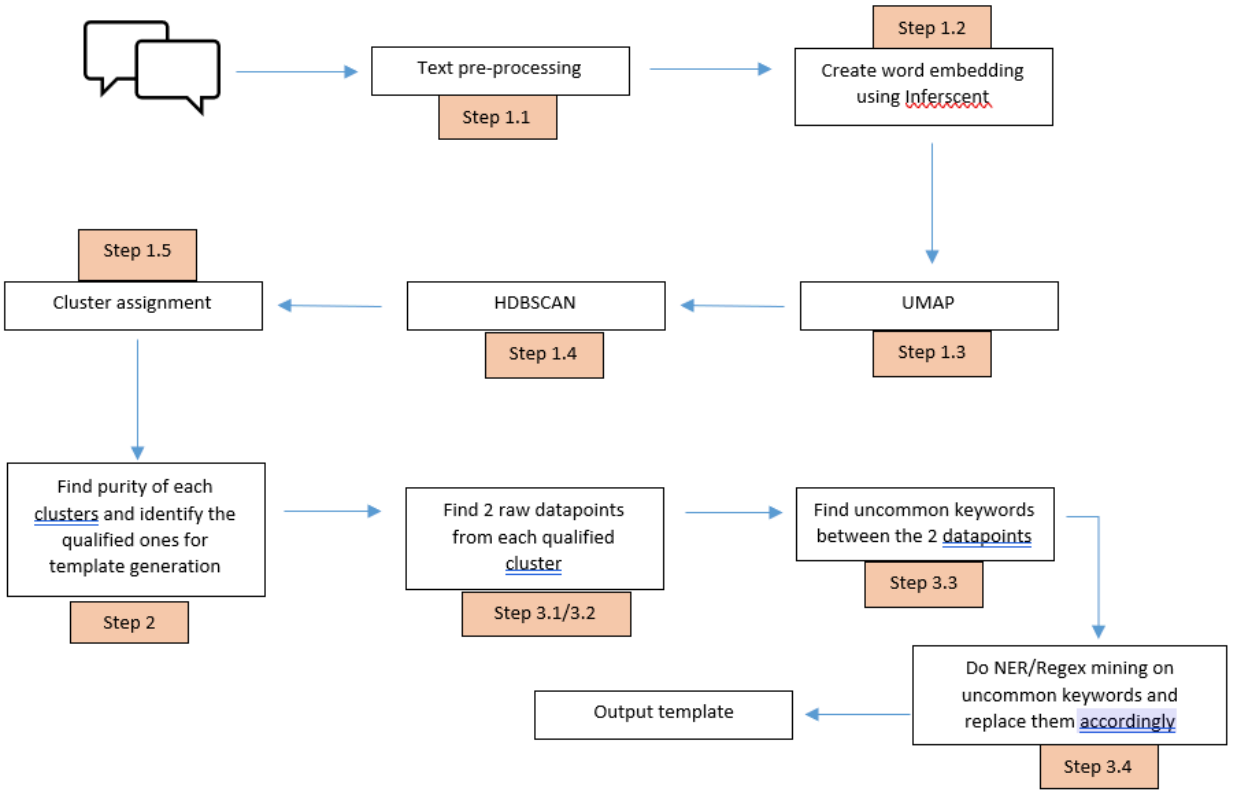
Template Generation:
Developing machine learning algorithms that leverage the extracted features to generate standardized email templates.

Incorporating adaptability to accommodate various communication contexts and user preferences.

Expected Contributions:

This research contributes to the field by introducing an innovative solution to the challenges posed by the manual effort involved in crafting emails. The automation of standardization and template generation is anticipated to reduce the cognitive load on users, enhance communication efficiency, and promote consistency in professional correspondence.

System architecture



Progress

1. Data extraction - We've extracted 1000 email samples from the logistics domain with a variety of communications.
2. Create embeddings - We've generated embeddings for each and every email by leveraging Hugging face Sentence_transoformers.
3. Dimensionality reduction - We've reduced the dimension of each embedding to a vector with size 100.
4. Clustering - We've created clusters from the embedding (after dimensionality reduction).
5. Written an algorithm to define whether a cluster is pure or not - When a cluster is centered around it's centroid/more dense around centroid, then it's pure else impure.
6. From impure clusters, it's hard to extract the templates.
7. From pure clusters, we're trying to extract the emails and standardizing them with NLP techniques.