

Papers on text summarization of twitter :

### 1. Twitter summarization with Social-Temporal Context :

- summarization framework T2ST –

Models temporal context from local and global perspectives mathematically

manually labelled Twitter dataset to evaluate the proposed framework

All the text summarisation papers that I looked at were not tackling the twitter feed summarisation in particular and the ones I found were more focussed on sentiment analysis and took larger text samples.

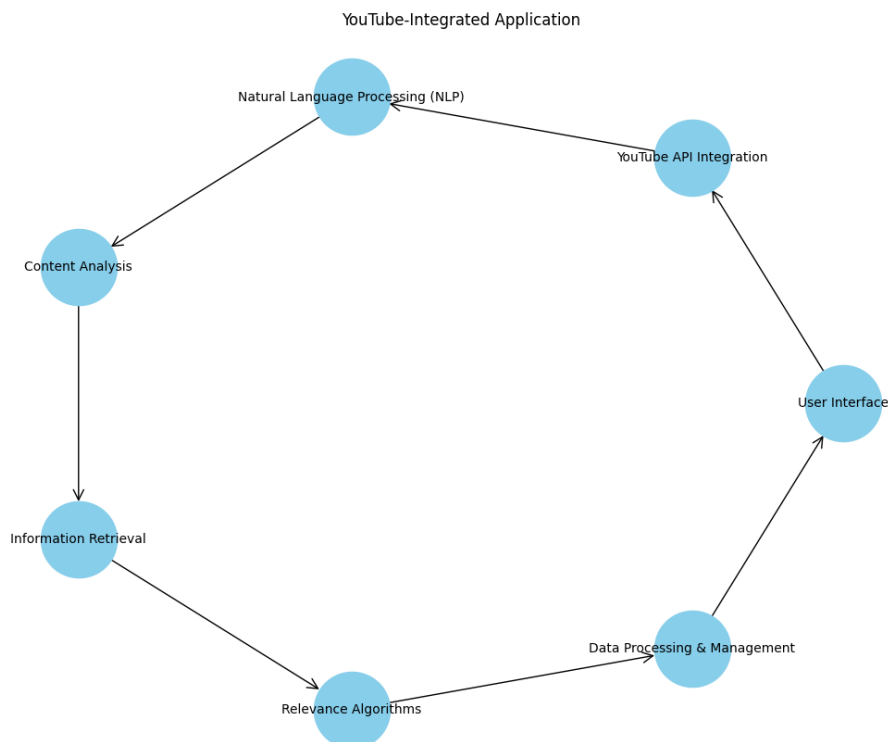
Considering the tweets on twitter have a character limit and are generally shorter, creation of a summary appears to be quite nuanced since agglomeration of similarly summarized texts will be difficult. Additionally It has been difficult to get data from twitter which is representative of my desired dataset and additionally since Twitter's takeover, The API access has become excessively more expensive which has made researching on Twitter more difficult.

Hence keeping this in mind, I have decided to scrap the idea of working on a text summarizer for Twitter and have decided to pursue the Youtube video summariser and LLM companion. The specifications along with the areas of focus are defined below.

I will be defining the problem in paper once for clarity's sake/

Given I wish to watch an hour long lecture by my professor on Youtube on Linear Algebra 2. Now I wish to ask questions on the very specific topic to the lecturer, Since that Question cannot be asked by the lecturer, I wish to equip an LLM with the capabilities to understand and work with the data provided in the video to understand the context and be able to answer to the query. Additionally if the query cannot be fulfilled then it looks into its training data to find the answer as governed by a Content analysis pipeline.

Looked at various papers – generated their own datasets in those papers. Not able to find relevant.



According to me the important things to focus on will be –

1. Contextual Understanding:

- NLP for user query
- Content analysis pipeline :
  - Method to analyse video content and everything around that.

2. Information Retrieval :

- Combining Multiple sources: - ask question -> answer from the video -> also add content and context from pretrained sources/ knowledge base. -> give a combined answer
- Relevance finding: A measure/ technique to determine relevance of external information to the context of the video

3. UI UX

4. Data Handling and Data Security.

What is required to be discussed:

1. How can we work with content analysis pipeline
2. Combining and working with 2 sources of info -> PreTrained info + Youtube Transcript data