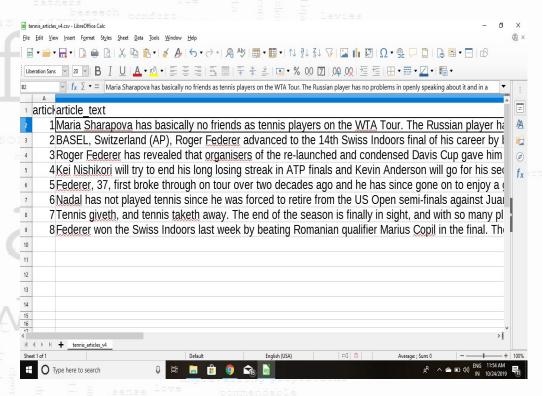


### **Problem Statement:**

To build a system that takes information from the user like an article or a document and summarizes the information provided into simple short sentences to avoid wastage of time and better productivity for users.

#### Abstract:

Firstly, we take a small test dataset which can be a document or an article. We have taken a sports news dataset which has all the sports content.



There are three main steps for summarizing documents. These are topic identification, interpretation and summary generation.

- Topic Identification: The most prominent information in the text is identified. There are different techniques for topic identification are used which are position, cue phrases, word frequency. Methods which are based on the position of phrases are the most useful methods for topic identification.
- Interpretation: Abstract summaries need to go through interpretation step. In this step, different subjects are fused in order to form a general content.
- Summary Generation: In this step, the system uses text generation method.

### Introduction:

- Text Summarization is one of the interesting applications of Machine Learning and Deep Learning.
- 2. It is one of the most challenging problems in the field of Deep Learning using **Natural Language Processing**.
- 3. It is a process of generating a concise and meaningful summary of text from multiple text resources such as books, news articles, blog posts, research papers and emails.
- 4. The demand for text summarization systems is spiking these days thanks to the availability of large amounts of textual data.
- 5. The intention is to create a coherent and fluent summary having only the main points outlined in the document.

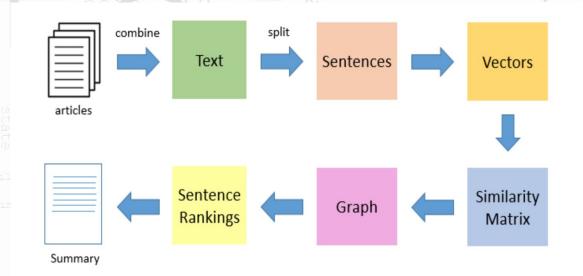
#### Research:

We have gone through some research papers to have a look at different approaches for summarizing single as well as multiple documents and articles.

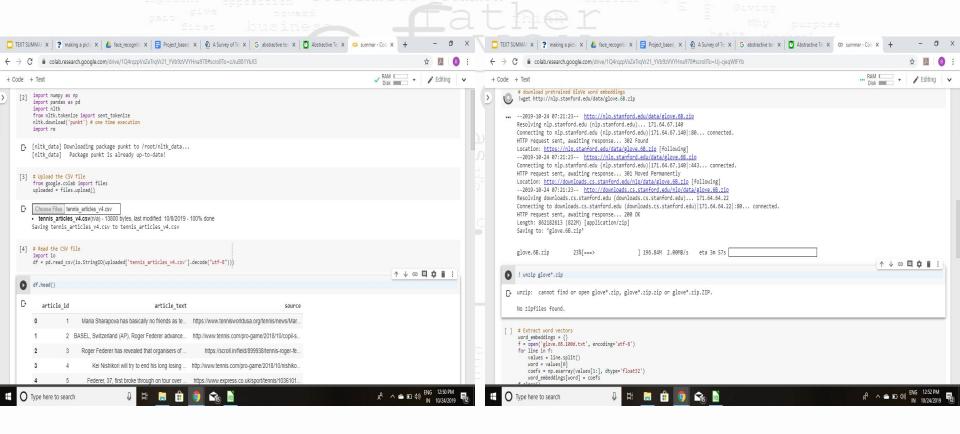
We have found abstractive summarization method to be a very powerful NLP technique to interpret text and generate new summary text.

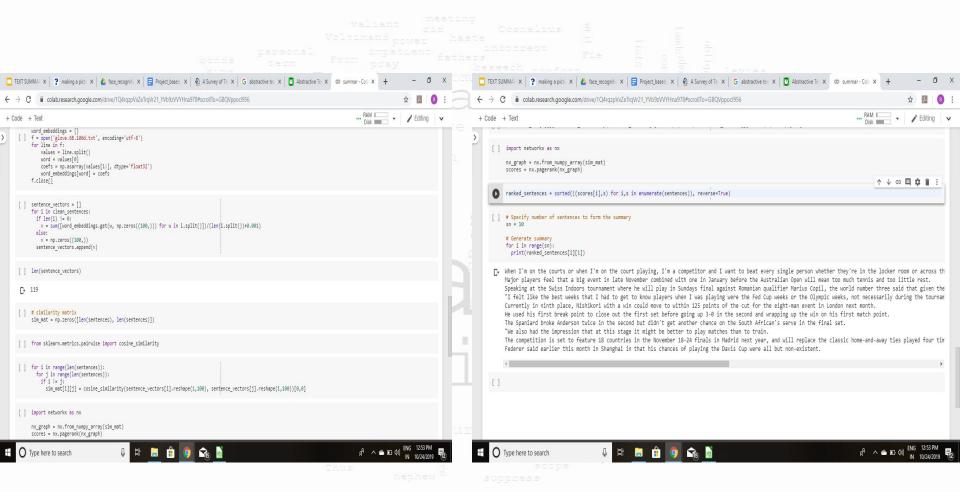
- This method first converts the paragraph into sentences then we remove all special characters from the sentences by text preprocessing.
- The sentence is then tokenized to get all the words that exist in a sentence.
- Next we need to find the weighted frequency of occurrences of all the words.

- The next step is to plug the weighted frequency in place of the corresponding words in original sentences and finding their sum.
- The final step is to sort the sentences in inverse order of their sum.
  The sentences give a pretty good summarization of what was said in the paragraph.



# Design of proposed system:





## **Further Approach:**

Currently, we have built a normal python code for summarizing single or multiple documents and articles.

We are planning to integrate this system with an Android Application i.e a user interface for the ease of users.

The user will be able to upload documents which they desire to summarize and then the application summaries the document. The user can save the summarized document on their mobile or laptop devices.

