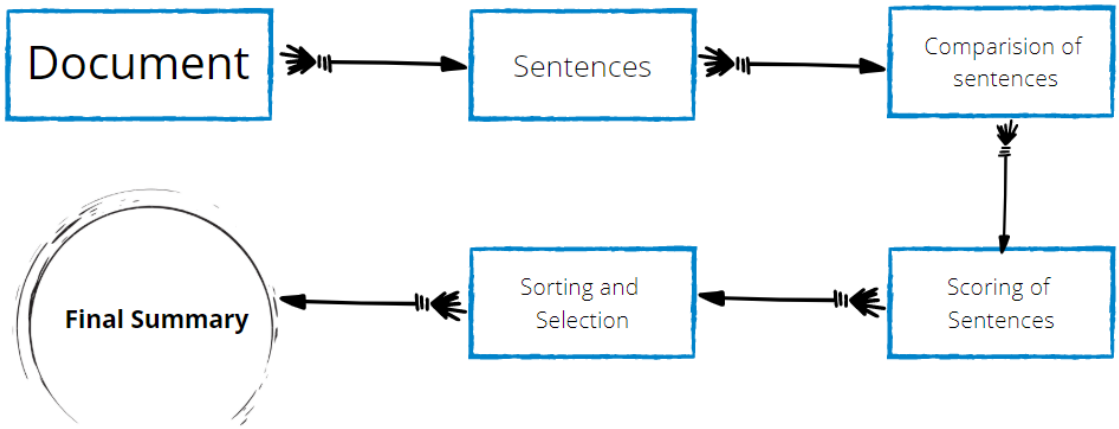
In Natural Language Processing, the summarization task can be broadly classified into two categories:

- Extractive Summarization -> Extractive summarization picks up sentences directly from the original document depending on their importance.
- Abstractive Summarization -> It tries to produce a bottom-up summary using sentences or verbal annotations that might not be a part of the original document.

We are using Extractive summarization for generating summary form the Text

Work Flow of Extractive summary



Result

```
  sorted_sentences: [
    [
      0.36734693877551017,
      'vedic priests will escort the prime minister while reciting mantras to the hanumangarhi temple'
    ],
    [
      0.26428571428571435,
      'prime minister narendra modi's first stop in uttar pradesh's ayodhya on august 5 will be a shrine dedicated to lord hanuman where he wi
ad of the grand foundation laying ceremony for the construction of the ram temple'
    ],
    [
      0.25925925925925936,
      'mahant raju das the head priest of hanumangarhi said they have been allocated seven minutes for prime minister's rituals at the temple
ency ani'
    ]
  ]
```

Implementation of Abstractive Summarization

Text Summarization Using Google T5 Model

T5 is a new transformer model given by Google that is trained in an end-to-end manner with text as input and modified text as output.

https://ai.googleblog.com/2020/02/exploring-transfer-learning-with-t5.html

Highlights

The T5 model, pre-trained on C4, achieves state-of-the-art results on many NLP benchmarks while being flexible enough to be fine-tuned to a variety of important downstream tasks.

C4 is a large dataset which was crawl through various news sources by google and due to its enormous and highly diverse it helps your model for giving more accurate results because the text from Wikipedia is high quality but uniform in style.
https://www.tensorflow.org/datasets/catalog/c4

Fine Tuning

We used the Inshorts dataset. Inshorts is a service that collects news from various sources and publishes them as a summary.

For training, we will consider the headline as the summary of a news article and the original article will be the context.

https://github.com/rojagtap/abstractive_summarizer

```
In [5]: news.head()
```

Out[5]:

	Headline	Short
0	4 ex-bank officials booked for cheating bank o...	The CBI on Saturday booked four former officia...
1	Supreme Court to go paperless in 6 months: CJI	Chief Justice JS Khehar has said the Supreme C...
2	At least 3 killed, 30 injured in blast in Sylh...	At least three people were killed, including a...
3	Why has Reliance been barred from trading in f...	Mukesh Ambani-led Reliance Industries (RIL) wa...
4	Was stopped from entering my own studio at Tim...	TV news anchor Arnab Goswami has said he was t...

```
In [6]: news.shape
Out[6]: (55104, 2)
```

Limitation of T5 model

1. It will generate word embedding of first 512 tokens(words).
2. For computing the result of 500 length article it will take 20 sec in cpu and 1.3 sec in gpu instance.

Other Existing Models for summarization

1. Bert Extractive Summarizer python library.
2. Pointer Generator Coverage Model Hybrid of Abstractive and Extractive Summarization.