



Text Summarization in NLP

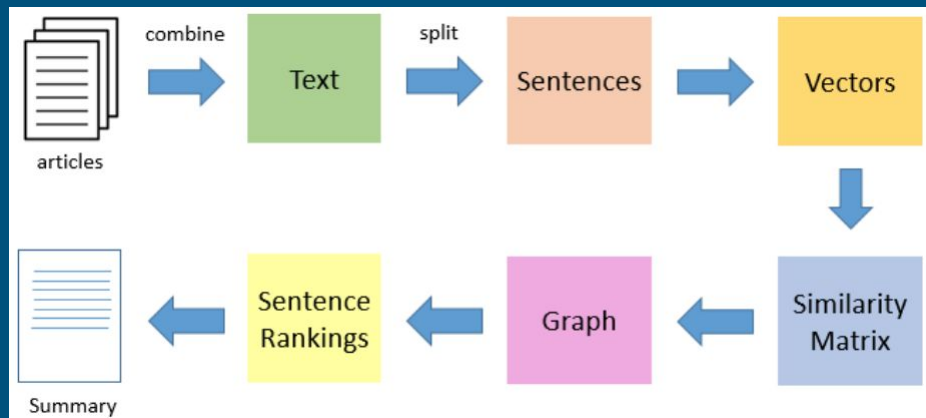
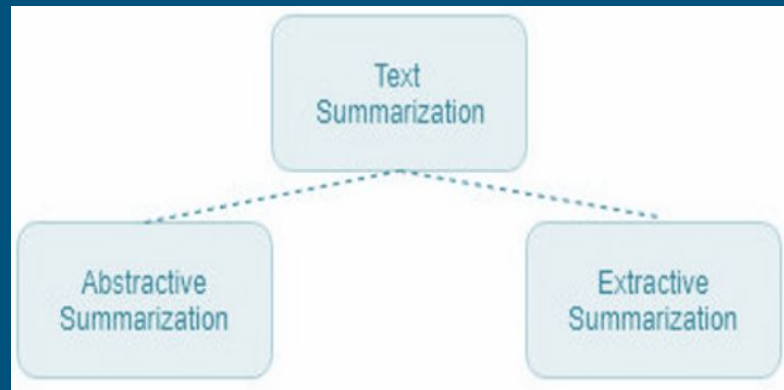


Zane Alderfer ~ IST 664



An Overview

Text summarization allows you to generate a small subset of a long text document. The idea is to find the crucial and essential information that summarizes the document effectively in a readable format. This allows the user to understand the makeup of a document in a faster manner to effectively utilize their time.



Why?

- Reduced reading time
- Improves effectiveness of indexing
- Less biased than human summarization
- Increases number of documents able to be processed
- Summaries make research document selection easier

Extractive vs. Abstractive

Extractive takes out important sentences or phrases and creates a summary from these excerpts. To deem what is important, a ranking algorithm is assigned to each sentence based on relevance to the document. The sentence ranking can be performed through processes like:

- Term frequency - inverse document frequency
- Graph-based methods
- Machine learning methods (SVM & Random Forests)

Abstractive generates new sentences based on the original document in a concise and easy-to-understand summary. This is done by a neural network model that understands the relationship between the words and phrases through natural language generation and deep learning.

Extractive vs. Abstractive

Input Article

Marseille, France (CNN) The French prosecutor leading an investigation into the crash of Germanwings Flight 9525 insisted Wednesday that he was not aware of any video footage from on board the plane. Marseille prosecutor Brice Robin told CNN that "so far no videos were used in the crash investigation." He added, "A person who has such a video needs to immediately give it to the investigators." Robin's comments follow claims by two magazines, German daily Bild and French Paris Match, of a cell phone video showing the harrowing final seconds from on board Germanwings Flight 9525 as it crashed into the French Alps. All 150 on board were killed. Paris Match and Bild reported that the video was recovered from a phone at the wreckage site. ...

Text Summarization Models

Abstractive summarization

Extractive summarization

Generated summary

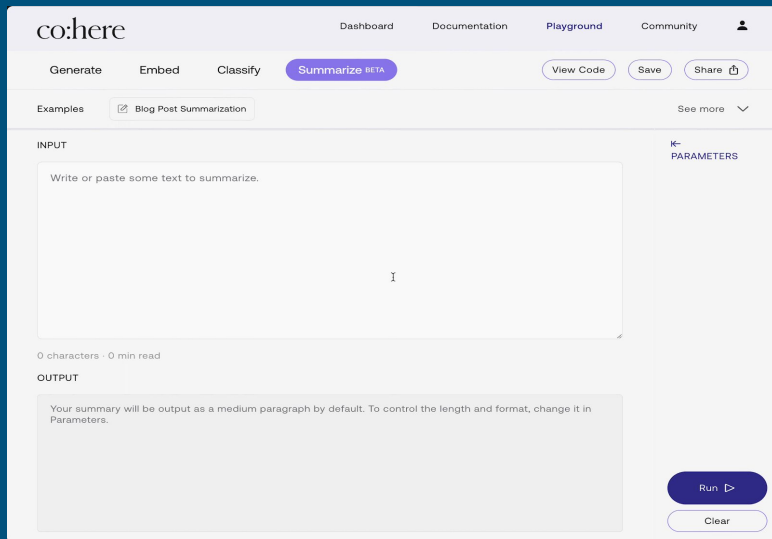
Prosecutor : " So far no videos were used in the crash investigation "

Extractive summary

marseille prosecutor brice robin told cnn that " so far no videos were used in the crash investigation . " robin \s comments follow claims by two magazines , german daily bild and french paris match , of a cell phone video showing the harrowing final seconds from on board germanwings flight 9525 as it crashed into the french alps . paris match and bild reported that the video was recovered from a phone at the wreckage site .

Cohere Summarization

A summarization API created by Cohere.ai that can deliver an extractive summarization or an abstractive summarization. Cohere playground for easy summarization use. Uses parameters such as length (length of summary: short, medium, long), format (paragraph or bullets), model (command, command-light, command-nightly), extractiveness (low, medium, high), temperature (0-5), and additional_command (free form instruction).



The screenshot shows the Cohere playground interface. At the top, there's a navigation bar with links to Dashboard, Documentation, Playground, and Community. Below this, there's a toolbar with buttons for Generate, Embed, Classify, and Summarize (which is highlighted in purple and labeled BETA). There are also buttons for View Code, Save, and Share. Under the Summarize button, there's a dropdown menu showing 'Blog Post Summarization' as the selected example. The main area is divided into two sections: INPUT and OUTPUT. The INPUT section has a large text area with a placeholder 'Write or paste some text to summarize.' and a character count '0 characters · 0 min read'. The OUTPUT section has a text area with a placeholder 'Your summary will be output as a medium paragraph by default. To control the length and format, change it in Parameters.' At the bottom right, there are buttons for 'Run' (with a play icon) and 'Clear'.

```
import cohere
co = cohere.Client('') # This is your trial API key
response = co.generate(
    model='command',
    prompt='Can you generate an abstractive summary from a research paper?
    Content of the research paper',
    max_tokens=2440,
    temperature=0.9,
    k=0,
    stop_sequences=[],
    return_likelihoods='NONE')
print('Prediction: {}'.format(response.generations[0].text))
```

<https://txt.cohere.com/summarize-beta/>

<https://blog.gopenai.com/humans-vs-coheres-summary-63e73a7b6c35>

Connexun Demo Example

Summarization API | Connexun

<https://www.wowessays.com/free-samples/good-culture-and-psychopathology-essay-example/>

References

<https://www.topcoder.com/thrive/articles/text-summarization-in-nlp>

<https://medium.com/analytics-vidhya/text-summarization-using-nlp-3e85ad0c6349>

<https://www.width.ai/post/4-long-text-summarization-methods>

<https://txt.cohere.com/summarize-beta/>

<https://www.analyticsvidhya.com/blog/2023/03/exploring-the-extractive-method-of-text-summarization/#:~:text=Extractive%20summarization%20techniques%20select%20and,of%20the%20original%20text%20intact.>

<https://www.edenai.co/post/best-summarization-apis>

<https://blog.gopenai.com/humans-vs-coheres-summary-63e73a7b6c35>

https://docs.cohere.com/reference/summarize-2?_gl=1*xxaoib*_gcl_au*MzQ4MzQ2NTY4LjE3MDEwMjM1MTM.