

Text Summarization Using Google T5 Model

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

- Extractive Summarization
- Abstractive Summarization

We are using Abstractive summarization for generating summary from the Text

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>

Text input

The US has "passed the peak" on new coronavirus cases, President Donald Trump said and predicted that some states would reopen this month. The US has over 637,000 confirmed Covid-19 cases and over 30,826 deaths, the highest for any country in the world. At the daily White House coronavirus briefing on Wednesday, Trump said new guidelines to reopen the country would be announced on Thursday after he speaks to governors. "We'll be the comeback kids, all of us," he said. "We want to get our country back." The Trump administration has previously fixed May 1 as a possible date to reopen the world's largest economy, but the president said some states may be able to return to normalcy earlier than that.

Summary from T5

The us has over 637,000 confirmed Covid-19 cases and over 30,826 deaths. President Donald Trump predicts some states will reopen the country in april, he said. "we'll be the comeback kids, all of us," the president says.

Analyzing the output:

If you see the algorithm has intelligently summarized with mentioning April although it was never mentioned in the original story. It also replaced he with the president. Cut short and removed the extra information after the comma in the sentence "over 30,826 deaths"