

Daily Log

Thursday January 09

Began to Look into Text Summarization Methods Mainly from Scratch. Found both

- <https://towardsdatascience.com/understand-text-summarization-and-create-your-own-summarizer-in-python-b26a9f09fc70>
- <https://blog.floydhub.com/gentle-introduction-to-text-summarization-in-machine-learning/>

Both Discussed the Basics of Text Summarization and Provided Small Pieces of Code to Get Started with Text Summarization and Packages Necessary to Download (Numpy, etc.)

Saturday January 11

Read <https://www.analyticsvidhya.com/blog/2019/06/comprehensive-guide-text-summarization-using-deep-learning-python/>. This Provided a Very Technical Understanding of Text Summarization and Discussed New Topics such as LSTMs (had to Look up Many New Ideas and Terms). Plan to Refer Back to this Site in the Following Week.

Timeline

Date	Goal	Met
Dec 13	Continue Implementing a Text Summarization Algorithm for a TED Talk	Yes, Got Better Results with New Implementation
Dec 20	Summarize a TED Talk by Implementing a Text-Summarization API	Yes, Was Able to Summarize a TED Talk to an Extent
Jan 10	Look Into Methods of Text Summarization from Scratch	Yes, Found a Few Sites About Text Summarization from Scratch
Jan 17	Continue to Look Into Methods of Text Summarization from Scratch	
Jan 24	Be Able to Summarize the Obama Speech in Less than 4 Sentences	

Reflection

I was glad to find two websites that provided a starting point for text summarization from scratch. The third site I looked at was very technical and thorough, but I believe if I can slowly understand the material, it will be very useful in creating my own text summarization script. Next week, I hope to continue to find more sites that discuss text summarization from scratch and maybe can even start implementing at the end of the week. In two weeks, I hope to be able to summarize the Obama speech in less than four sentences, something the text summarization API struggled to do. After that, I plan to look into the state of the art for audio-to-text and hope to find one that can deal with the punctuation issue I ran into.