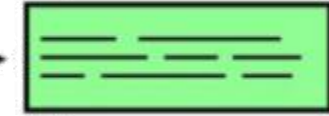


Digital Assignment - 3

Full
Article



NLP



Extracted
Summary

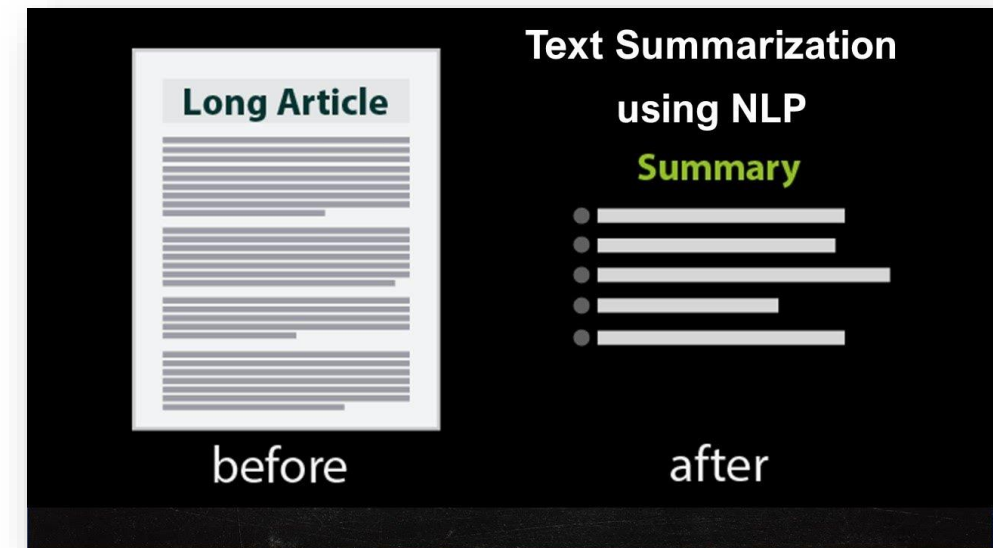
Topic: - Text Summarization

Sahil Sachin Donde - 19BCE1353

CSE4022 - Natural Language Processing

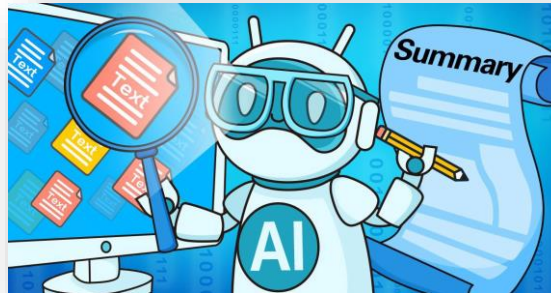
Faculty: - Dr. Vani V

B1 + TB1 Slot



Introduction to Text Summarization in NLP

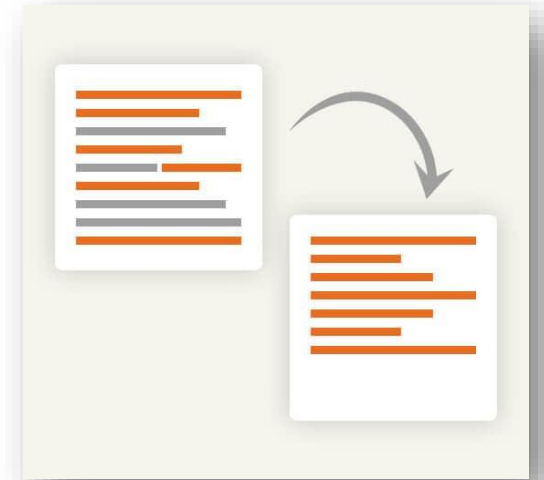
- *Text summarization* - process of summarizing the information of large texts
- Glancing short news summary/articles, then read more details if interested
- It is possible to create summaries automatically
- Extracting summaries from original text w/o losing vital information - Text Summarization
- Summary should be **fluent**, continuous and depicts the **significant** info
- Inshorts app & various other news aggregators uses text summarization algorithms.



Types of Text Summarization

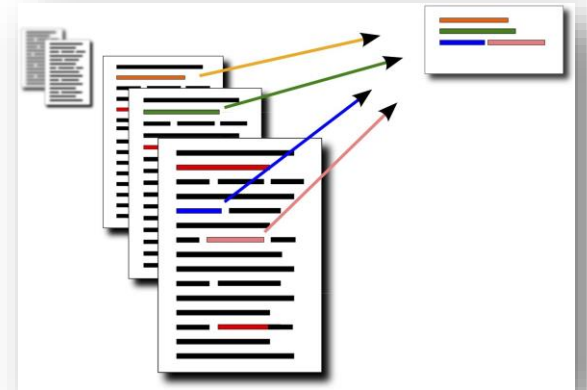
Extractive Text Summarization

- 1st Traditional method developed
- Identifies significant sentences of text
- Summary obtained contains exact sentences from input text



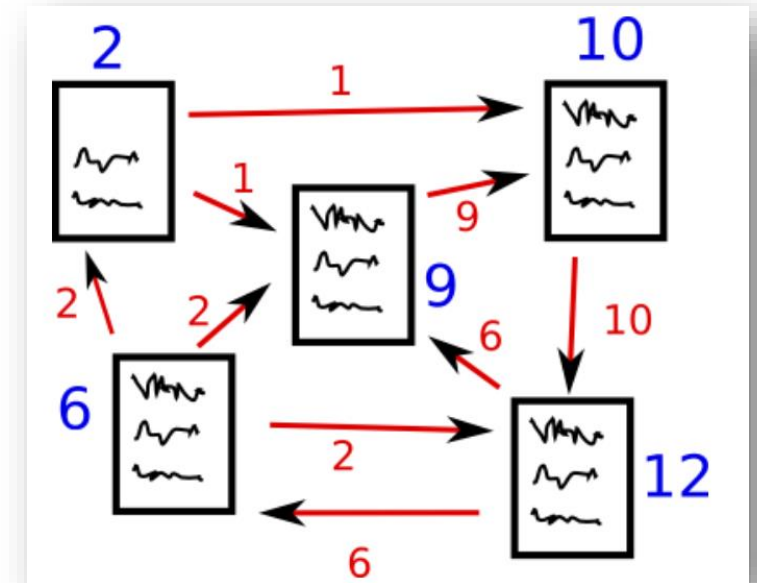
Abstractive Text Summarization

- Identifies important sections, interpret the context.
- Core information is conveyed through shortest text possible
- Sentences in gist are generated, not just extracted as such



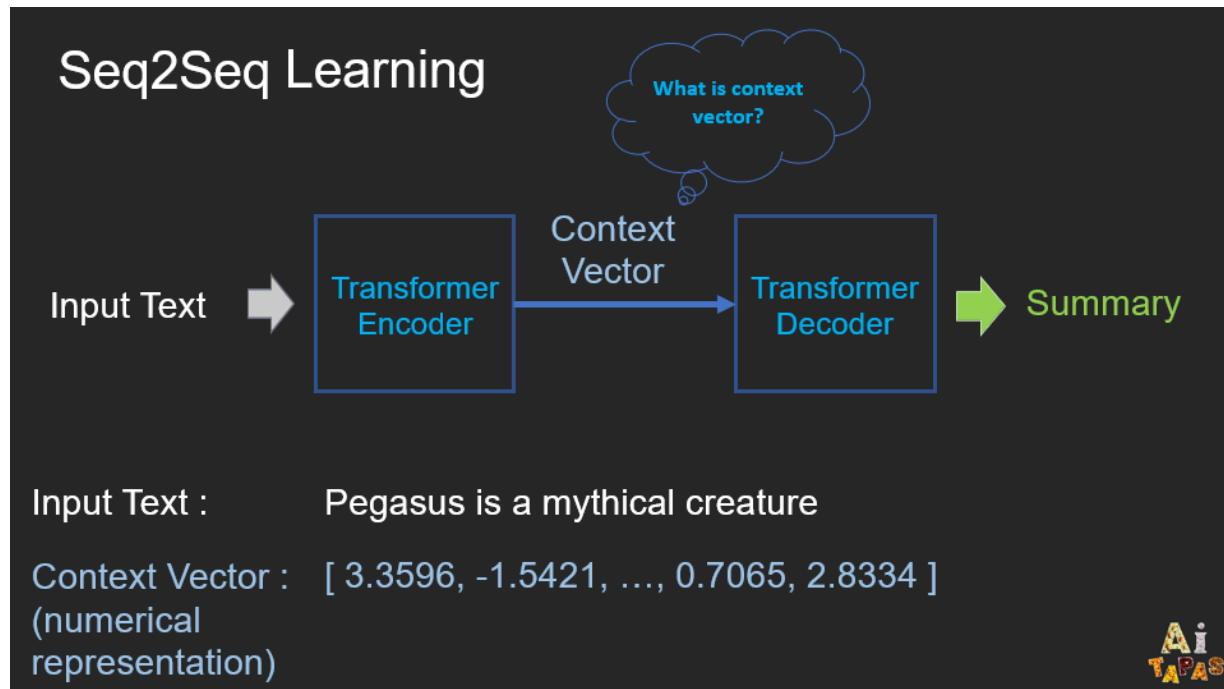
Extractive Text Summarization using Gensim – Text Rank

- Gensim - handy python library for performing NLP tasks
- The text summarization using Gensim library is based on TextRank Algorithm
- TextRank is an extractive summarization technique
- Words with more frequency are significant
- Sentences containing highly frequent words are important
- Algorithm assigns scores to each sentence in the text
- Top-ranked sentences make it to the summary



Abstractive Text Summarization using – PEGASUS Model

- PEGASUS uses an encoder-decoder model for sequence-to-sequence learning
- Encoder understands the context of the whole input text and encodes it a context vector
- Context Vector – a numerical representation of the input text



- Context vector - fed to decoder
- Decoder decodes context vector to produce gist
- Pegasus automates the selection of “important” sentences by perform pre-training of the model on a large corpus, i.e., 350 million web pages and 1.5 billion news articles.

References

- <https://neurondai.medium.com/automatic-text-summarization-system-using-transformers-are-you-tired-of-reading-a-long-paper-22d2cd9f5260>
- <https://towardsdatascience.com/how-to-perform-abstractive-summarization-with-pegasus-3dd74e48bafb>

Thank You